six sigma

advanced tools for black belts and master black belts

Loon Ching Tang | Thong Ngee Goh | Hong See Yam | Timothy Yoap

WILEY

Six Sigma

Advanced Tools for Black Belts and Master Black Belts

Loon Ching Tang
National University of Singapore, Singapore

Thong Ngee Goh
National University of Singapore, Singapore

Hong See Yam
Seagate Technology International, Singapore

Timothy Yoap
Flextronics International, Singapore

John Wiley & Sons, Ltd

Six Sigma

Six Sigma

Advanced Tools for Black Belts and Master Black Belts

Loon Ching Tang
National University of Singapore, Singapore

Thong Ngee Goh
National University of Singapore, Singapore

Hong See Yam
Seagate Technology International, Singapore

Timothy Yoap
Flextronics International, Singapore

John Wiley & Sons, Ltd

Copyright © 2006 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, ONT, L5R 4J3

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Six sigma: advanced tools for black belts and master black belts/Loon Ching Tang . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-470-02583-3 (cloth : alk. paper)
ISBN-10: 0-470-02583-2 (cloth : alk. paper)
1. Six sigma (Quality control standard) 2. Total quality management. I. Tang, Loon Ching.
TS156.S537 2006
658.5′62–dc22

2006023985

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-02583-3 (HB)
ISBN-10 0-470-02583-2 (HB)

Typeset in 10/12pt BookAntiqua by TechBooks, New Delhi, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

Contents

Preface ..... **xi**

PART A: SIX SIGMA: PAST, PRESENT AND FUTURE

1 Six Sigma: A Preamble **3**
H. S. Yam

1.1 Introduction 3
1.2 Six Sigma Roadmap: DMAIC 4
1.3 Six Sigma Organization 7
1.4 Six Sigma Training 8
1.5 Six Sigma Projects 10
1.6 Conclusion 17
References 17

2 A Strategic Assessment of Six Sigma **19**
T. N. Goh

2.1 Introduction 19
2.2 Six Sigma Framework 20
2.3 Six Sigma Features 21
2.4 Six Sigma: Contrasts and Potential 22
2.5 Six Sigma: Inherent Limitations 23
2.6 Six Sigma in the Knowledge Economy 25
2.7 Six Sigma: Improving the Paradigm 27
References 28

3 Six Sigma SWOT **31**
T. N. Goh and L. C. Tang

3.1 Introduction 31
3.2 Outline of Six Sigma 32
3.3 SWOT Analysis of Six Sigma 32
3.4 Further Thoughts 37
References 39

4 The Essence of Design for Six Sigma 41
L. C. Tang

4.1 Introduction 41
4.2 The IDOV Roadmap 42
4.3 The Future 48
References 48

5 Fortifying Six Sigma with OR/MS Tools 49
L. C. Tang, T. N. Goh and S. W. Lam

5.1 Introduction 49
5.2 Integration of OR/MS into Six Sigma Deployment 50
5.3 A New Roadmap for Six Sigma Black Belt Training 52
5.4 Case Study: Manpower Resource Planning 58
5.5 Conclusions 68
References 68

PART B: MEASURE PHASE

6 Process Variations and Their Estimates 73
L. C. Tang and H. S. Yam

6.1 Introduction 73
6.2 Process Variability 76
6.3 Nested Design 79
References 83

7 Fishbone Diagrams vs. Mind Maps 85
Timothy Yoap

7.1 Introduction 85
7.2 The Mind Map Step by Step 86
7.3 Comparison between Fishbone Diagrams and Mind Maps 87
7.4 Conclusion and Recommendations 91
References 91

8 Current and Future Reality Trees 93
Timothy Yoap

8.1 Introduction 93
8.2 Current Reality Tree 94
8.3 Future Reality Tree (FRT) 97
8.4 Comparison with Current Six Sigma Tools 101
8.5 Conclusion and Recommendations 105
References 105

9 Computing Process Capability Indices for Nonnormal Data: A Review and Comparative Study 107
L. C. Tang, S. E. Than and B. W. Ang

9.1 Introduction 107
9.2 Surrogate PCIs for Nonnormal Data 108

9.3 Simulation Study 113
9.4 Discussion of Simulation Results 127
9.5 Conclusion 128
References 129

10 Process Capability Analysis for Non-Normal Data with MINITAB 131
Timothy Yoap

10.1 Introduction 131
10.2 Illustration of the Two Methodologies Using a Case Study Data Set ... 132
10.3 A Further Case Study 141
10.4 Monte Carlo Simulation 145
10.5 Summary 149
References 149

PART C: ANALYZE PHASE

11 Goodness-of-Fit Tests for Normality 153
L. C. Tang and S. W. Lam

11.1 Introduction 153
11.2 Underlying Principles of Goodness-of-Fit Tests 154
11.3 Pearson Chi-Square Test 155
11.4 Empirical Distribution Function Based Approaches 157
11.5 Regression-Based Approaches 163
11.6 Fisher's Cumulant Tests 167
11.7 Conclusion 170
References 170

12 Introduction to the Analysis of Categorical Data 171
L.C. Tang and S. W. Lam

12.1 Introduction 171
12.2 Contingency Table Approach 173
12.3 Case Study 176
12.4 Logistic Regression Approach 181
12.5 Conclusion 193
References 193

13 A Graphical Approach to Obtaining Confidence Limits of C_{pk} 195
L. C. Tang, S. E. Than and B. W. Ang

13.1 Introduction 196
13.2 Graphing C_p, k *and* p 197
13.3 Confidence Limits for k 199
13.4 Confidence Limits For C_{pk} 201
13.5 A Simulation Study 203
13.6 Illustrative Examples 206
13.7 Comparison with Bootstrap Confidence Limits 207
13.8 Conclusions 209
References 210

14 Data Transformation for Geometrically Distributed Quality Characteristics 211
T. N. Goh, M. Xie and X. Y. Tang

14.1 Introduction 211
14.2 Problems of Three-Sigma Limits for the G Chart 212
14.3 Some Possible Transformations 213
14.4 Some Numerical Comparisons 216
14.5 Sensitivity Analysis of the Q Transformation 219
14.6 Discussion 221
References 221

15 Development of A Moisture Soak Model For Surface Mounted Devices 223
L. C. Tang and S. H. Ong

15.1 Introduction 223
15.2 Experimental Procedure and Results 225
15.3 Moisture Soak Model 227
15.4 Discussion 234
References 236

PART D: IMPROVE PHASE

16 A Glossary for Design of Experiments with Examples 239
H. S. Yam

16.1 Factorial Designs 239
16.2 Analysis of Factorial Designs 242
16.3 Residual Analysis 243
16.4 Types of Factorial Experiments 244
16.5 Fractional Factorial Designs 246
16.6 Robust Design 250

17 Some Strategies for Experimentation under Operational Constraints 257
T. N. Goh

17.1 Introduction 257
17.2 Handling Insufficient Data 258
17.3 Infeasible Conditions 258
17.4 Variants of Taguchi Orthogonal Arrays 260
17.5 Incomplete Experimental Data 262
17.6 Accuracy of Lean Design Analysis 262
17.7 A Numerical Illustration 263
17.8 Concluding Remarks 264
References 265

18 Taguchi Methods: Some Technical, Cultural and Pedagogical Perspectives 267
T. N. Goh

18.1 Introduction 268

18.2 General Approaches to Quality 268
18.3 Stages in Statistical Applications 269
18.4 The Taguchi Approach 272
18.5 Taguchi's 'Statistical Engineering' 273
18.6 Cultural Insights 282
18.7 Training and Learning 286
18.8 Concluding Remarks 291
18.9 Epilogue 292
References 293

19 Economical Experimentation via 'Lean Design' 297
T. N. Goh

19.1 Introduction 297
19.2 Two Established Approaches 298
19.3 Rationale of Lean Design 298
19.4 Potential of Lean Design 299
19.5 Illustrative Example 302
19.6 Possible Applications 303
19.7 Concluding Remarks 305
References 306

20 A Unified Approach for Dual Response Surface Optimization 307
L. C. Tang and K. Xu

20.1 Introduction 307
20.2 Review of Existing Techniques for Dual Response Surface Optimization 308
20.3 Example 1 314
20.4 Example 2 319
20.5 Conclusions 320
References 322

PART E: CONTROL PHASE

21 Establishing Cumulative Conformance Count Charts 325
L. C. Tang and W. T. Cheong

21.1 Introduction 325
21.2 Basic Properties of the CCC Chart 326
21.3 CCC Scheme with Estimated Parameter 327
21.4 Constructing A CCC Chart 330
21.5 Numerical Examples 336
21.6 Conclusion 339
References 340

22 Simultaneous Monitoring of the Mean, Variance and Autocorrelation Structure of Serially Correlated Processes 343
O. O. Atienza and L. C. Tang

22.1 Introduction 344

22.2 The Proposed Approach 345
22.3 ARL Performance 346
22.4 Numerical Example 349
22.5 Conclusion 351
References 352

23 Statistical Process Control for Autocorrelated Processes : A Survey and An Innovative Approach 353
L. C. Tang and O. O. Atienza

23.1 Introduction 353
23.2 Detecting Outliers and Level Shifts 355
23.3 Behavior of $\lambda_{LS,t}$ 358
23.4 Proposed Monitoring Procedure 363
23.5 Conclusions 366
References 368

24 Cumulative Sum Charts with Fast Initial Response 371
L. C. Tang and O. O. Atienza

24.1 Introduction 371
24.2 Fast Initial Response 374
24.3 Conclusions 379
References 379

25 CUSUM and Backward CUSUM for Autocorrelated Observations 381
L. C. Tang and O. O. Atienza

25.1 Introduction 381
25.2 Backward CUSUM 382
25.3 Symmetric Cumulative Sum Schemes 387
25.4 CUSUM Scheme for Autocorrelated Observations 391
25.5 Conclusion 404
References 405

Index 407

Preface

The only place where Quality comes before Statistics is in the dictionary.
(T. N. Goh)

Six Sigma has come a long way since its introduction in the mid-1980s. Our association with the subject began in the 1990s when a number of multinational corporations in Singapore began to deploy Six Sigma in pursuit of business excellence. Prior to this, some of us had been working on statistical quality improvement techniques for more than two decades. It was apparent at the outset that the strength of Six Sigma is not in introducing new statistical techniques as it relies on well-established and proven tools; Six Sigma derives its power from the way corporate mindsets are changed towards the application of statistical tools, from top business leaders to those on the production floor. We are privileged to be part of this force for change through our involvement in Six Sigma programs with many companies in the Asia-Pacific region.

Over the last decade, as Six Sigma has taken root in a number of corporations in the region, the limitations of existing tools have surfaced and the demand for innovative solutions has increased. This has coincided with the rapid evolution of Six Sigma as it permeated across various industries, and in many cases the conventional Six Sigma toolset is no longer sufficient to provide adequate solutions. This has opened up many research opportunities and motivated close collaborations between academia and industrial practitioners. This book represents part of this effort to bring together practitioners and academics to work towards the common goal of providing an advanced reference for Six Sigma professionals, particularly Black Belts and Master Black Belts.

The book is organized into five parts, of five chapters each. Each of the parts represents respectively the define, measure, analyze, improve and control phases of the traditional Six Sigma roadmap. Part A presents a strategic assessment of Six Sigma and its SWOT analysis, followed by discussions of current interests in Six Sigma, including Design for Six Sigma as well as a new improvement roadmap for transactional Six Sigma.

In Part B, basic concepts of variability and some useful qualitative tools such as mind maps and reality trees are presented. Capability analysis for non-normal data is also discussed in two chapters focusing respectively on the theoretical and practical aspects.

In Part C, we start with a chapter reviewing goodness-of-fit tests for normality, and then give a basic treatment of categorical data. These techniques are instrumental in analyzing industrial data. A novel graphical approach in determining the confidence interval for the process capability index, C_{pk}, is then presented. This is followed by an examination of the transformation of geometrically distributed variables. These two chapters are based on material previously published in the journal *Quality and Reliability Engineering International*. A case study to illustrate how to do subset selection in multiple regression is given and could serve as an application guide.

Part D begins with a glossary list in design of experiment (DOE) and is based on four previously published papers by the authors. These papers aim to illustrate important concepts and methodology in DOE in a way that is appealing to Six Sigma practitioners.

Finally, in Part E, some advanced charting techniques are presented. These include the cumulative conformance count chart, cumulative sum (CUSUM) charts with head-start features, and CUSUM charts for autocorrelated processes. Particular emphasis is placed on the implementation of statistical control for autocorrelated processes which are quite common in today's industry with automatic data loggers. Notably, we include a contributed paper by Dr Orlando Atienza that proposes a novel approach to monitoring changes in mean, variance and autocorrelation structure simultaneously.

This book is a collection of concepts and selected tools that are important to the mature application of the Six Sigma methodology. Most of them are motivated by questions asked by students, trainees and colleagues over the last decade in the course of our training and consulting activities in industry. Some of these have been presented to graduate students to get their research work off the ground. We are thus indebted to many people who have contributed in one way or another to the development of the material, and it is not easy to mention every one of them. In particular, our colleagues and students at the National University of Singapore and many Master Black Belts, Black Belts, and Green Belts of Seagate Technology have been our sources of inspiration. We would also like to thank Dr W. T. Cheong (now with Intel) and Mr Tony Halim who have assisted in the preparation of the manuscript.

L. C. Tang
T. N. Goh
H. S. Yam
T. Yoap
Singapore, April 2006

Part A

Six Sigma: Past, Present and Future

1

Six Sigma: A Preamble

H. S. Yam

Six Sigma is a rigorous and highly disciplined business process adopted by companies to help focus on developing and delivering robust, near-perfect products and services. In this opening chapter, we first present the underlying motivation for Six Sigma. While Six Sigma has demonstrated itself to be of much value in manufacturing operations, its full potential is not realized till it has been proliferated and leveraged across the multitude of functions in a business entity. To achieve this end, a well-defined vision and roadmap, along with structured roles, are necessary. In this chapter, we present a brief description of the DMAIC roadmap and the organizational structure in a typical Six Sigma deployment. This is followed by a discussion of how to customize appropriate levels of Six Sigma training for these various roles. Finally, an example of a Six Sigma project is presented to illustrate the power of integrating existing technical expertise/knowledge with the Six Sigma methodology and tools in resolving leveraged problems.

1.1 INTRODUCTION

Six Sigma has captured the attention of chief executive officers (CEOs) from multi-billion corporations and financial analysts on Wall Street over the last decade. But what is it?

Mikel Harry, president and CEO of Six Sigma Academy Inc, defines it as 'a business process that allows companies to drastically improve their bottom line by designing and monitoring everyday business activities in ways that minimize waste and resources while increasing customer satisfaction'.[1] Pande *et al.* call it 'a comprehensive and flexible system for achieving, sustaining and maximizing business success, ... uniquely driven by close understanding of customer needs, disciplined use of facts, data and statistical analysis, with diligent attention to managing, improving

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

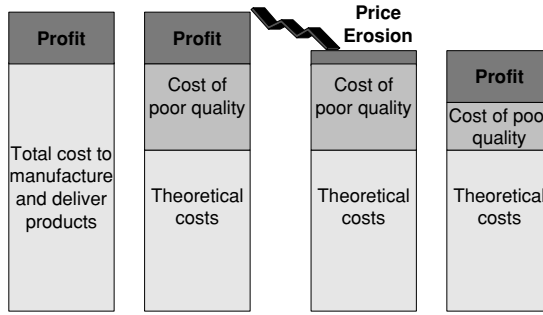


Figure 1.1 Relationship between price erosion, cost of poor quality and profit.

and reinventing business processes'.[2] Contrary to general belief, the goal of Six Sigma is not to achieve 6σ levels of quality (i.e. 3.4 defects per million opportunities). It is about improving profitability; improved quality and efficiency are the immediate by-products.[1]

Some have mistaken Six Sigma as another name for *total quality management* (TQM). In TQM, the emphasis is on the involvement of those closest to the process, resulting in the formation of *ad hoc* and self-directed improvement teams. Its execution is owned by the quality department, making it difficult to integrate throughout the business. In contrast, Six Sigma is a business strategy supported by a quality improvement strategy.[3] While TQM, in general, sets vague goals of customer satisfaction and highest quality at the lowest price, Six Sigma focuses on bottom-line expense reductions with measurable and documented results. Six Sigma is a strategic business improvement approach that seeks to increase both customer satisfaction and a company's financial health.[4]

Why should any business consider implementing Six Sigma? Today, there is hardly any product that can maintain a monopoly for long. Hence, price erosion in products and services is inherent. Profit is the difference between revenues and the cost of manufacturing (or provision of service), which in turn comprises the theoretical cost of manufacturing (or service) and the hidden costs of poor quality (Figure 1.1). Unless the cost component is reduced, price erosion can only bite into our profits, thereby reducing our long-term survivability. Six Sigma seeks to improve bottom-line profits by reducing the hidden costs of poor quality.

The immediate benefits enjoyed by businesses implementing Six Sigma include operational cost reduction, productivity improvement, market-share growth, customer retention, cycle-time reduction and defect rate reduction.

1.2 SIX SIGMA ROADMAP: DMAIC

In the early phases of implementation in a manufacturing environment, Six Sigma is typically applied in manufacturing operations, involving personnel mainly from process and equipment engineering, manufacturing and quality departments. For Six Sigma to be truly successful in a manufacturing organization, it has to be proliferated

across its various functions – from design engineering, through materials and shipping, to sales and marketing, and must include participation from supporting functions such as information technology, human resources and finance. In fact, there is not a single function that can remain unaffected by Six Sigma. However, widespread proliferation would not be possible without appropriate leadership, direction and collaboration.

Six Sigma begins by identifying the needs of the customer. These needs generally fall under the categories of timely delivery, competitive pricing and zero-defect quality. The customer's needs are then internalized as performance metrics (e.g. cycle time, operational costs and defect rate) for a Six Sigma practicing company. Target performance levels are established, and the company then seeks to perform around these targets with minimal variation.

For successful implementation of Six Sigma, the business objectives defined by top-level executives (such as improving market share, increasing profitability, and ensuring long-term viability) are passed down to the operational managers (such as yield improvement, elimination of the 'hidden factory' of rework, and reduction in labor and material costs). From these objectives, the relevant processes are targeted for defect reduction and process capability improvement.

While conventional improvement programs focus on improvements to address the defects in the 'output', Six Sigma focuses on the process that creates or eliminates the defects, and seeks to reduce variability in a process by means of a systematic approach called the *breakthrough strategy*, more commonly known as the *DMAIC methodology.* DMAIC is an acronym for Define–Measure–Analyze–Improve–Control, the various development phases for a typical Six Sigma project.

The *define* phase sets the stage for a successful Six Sigma project by addressing the following questions:

- What is the problem to be addressed?
- What is the goal? And by when?
- Who is the customer impacted?
- What are the CTQs in-concern?
- What is the process under investigation?

The *measure* phase serves to validate or redefine the problem. It is also the phase where the search for root causes begins by addressing:

- the focus and extent of the problem, based on measures of the process;
- the key data required to narrow down the problem to its major factors or vital few root causes.

In the *analyze* phase, practical business or operational problems are turned into statistical problems (Figure 1.2). Appropriate statistical methods are then employed:

- to discover what we do not know (exploratory analysis);
- to prove/disprove what we suspect (inferential analysis).

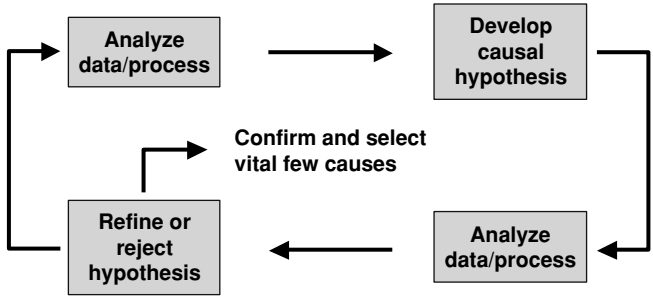


Figure 1.2 The analyze phase.

The *improve* phase focuses on discovering the key variables (inputs) that cause the problem. It then seeks to address the following questions:

- What possible actions or ideas are required to address the root cause of the problem and to achieve the goal?
- Which of the ideas are workable potential solutions?
- Which solution is most to likely achieve the desired goal with the least cost or disruption?
- How can the chosen solution be tested for effectiveness? How can it be implemented permanently?

In the *control* phase, actions are established to ensure that the process is monitored continuously to facilitate consistency in quality of the product or service (Figure 1.3). Ownership of the project is finally transferred to a finance partner who will track the financial benefits for a specified period, typically 12 months.

In short, the DMAIC methodology is a disciplined procedure involving rigorous data gathering and statistical analysis to identify sources of errors, and then seeking for ways to eliminate these causes.

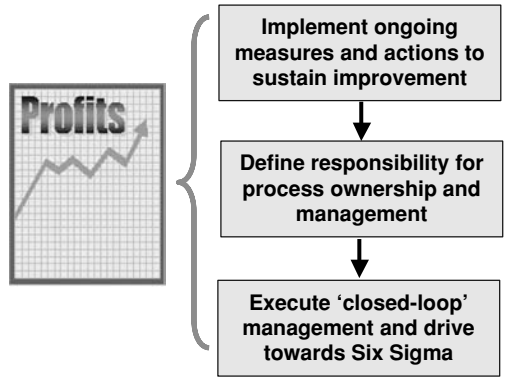

Profits
Implement ongoing measures and actions to sustain improvement
Define responsibility for process ownership and management
Execute 'closed-loop' management and drive towards Six Sigma

Figure 1.3 Six Sigma culture drives profitability.

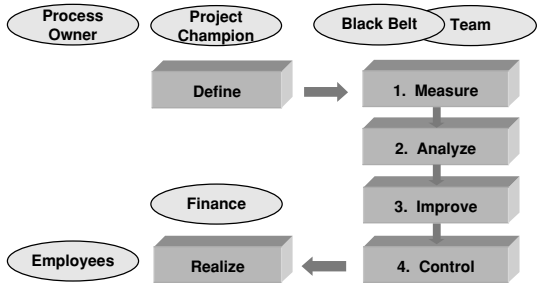


Figure 1.4 Interactions of stakeholders in various phases of a Six Sigma project.

1.3 SIX SIGMA ORGANIZATION

For best results, the DMAIC methodology must be combined with the right people (Figure 1.4). At the center of all activities is the Black Belt, an individual who works full-time on executing Six Sigma projects. The Black Belt acts as the project leader, and is supported by team members representing the functional groups relevant to the project. The Champion, typically a senior manager or director, is both sponsor and facilitator to the project and team. The Process Owner is the manager who receives the handoff from the team, and is responsible for implementation and maintenance of the agreed solution. The Master Black Belt is the consultant who provides expert advice and assistance to the Process Owner and Six Sigma teams, in areas ranging from statistics to change management to process design strategies.

Contrary to general belief, the success of Six Sigma does not lie in the hands of a handful of Black Belts, led by a couple of Master Black Belts and Champions. To realize the power of Six Sigma, a structure of roles and responsibilities is necessary (Figure 1.5). As Six Sigma is targeted at improving the bottom-line performance of a company, its support must stem from the highest levels of *executive management*. Without an overview of the business outlook and an understanding of the company's strengths and weaknesses, deployment of Black Belts to meet established corporate-level goals and targets within an expected time frame would not be possible.

The *Senior Champion* is a strong representative from the executive group and is accountable to the company's president. He/she is responsible for the day-to-day

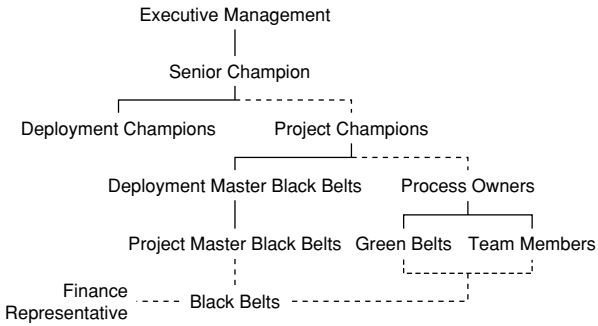


Figure 1.5 The reporting hierarchy of the Six Sigma team.

corporate-level management of Six Sigma, as well as obtaining the business unit executives to commit to specific performance targets and financial goals.

The *Deployment Champions* are business unit directors responsible for the development and execution of Six Sigma implementation and deployment plans for their defined respective areas of responsibility. They are also responsible for the effectiveness and efficiency of the Six Sigma support systems. They report to the Senior Champion, as well as the executive for their business unit.

The *Project Champions* are responsible for the identification, selection, execution and follow-on of Six Sigma projects. As functional and hierarchical managers of the Black Belts, they are also responsible for their identification, selection, supervision and career development.

The *Deployment Master Black Belts* are responsible for the long-range technical vision of Six Sigma and the development of its technology roadmaps, identifying and transferring new and advanced methods, procedures and tools to meet the needs of the company's diverse projects.

The *Project Master Black Belts* are the technical experts responsible for the transfer of Six Sigma knowledge, either in the form of classroom training or on-the-job mentoring. It is not uncommon to find some Project Master Black Belts doubling up as Deployment Master Black Belts.

The *Black Belts* play the lead role in Six Sigma, and are responsible for executing application projects and realizing the targeted benefits. Black Belts are selected for possession of both hard technical skills and soft leadership skills, as they are also expected to work with, mentor and advise middle management on the implementation of process-improvement plans. At times, some may even be leading cross-functional and/or cross-site projects. While many companies adopt a 2-year conscription for their Black Belts, some may chose to offer the Black Belt post as a career.

The *Process Owners* are the line managers of specific business processes who review the recommendations of the Black Belts, and ensure that process improvements are captured and sustained through their implementation and/or compliance.

Green Belts may be assigned to assist in one or more Black Belts projects, or they may be leaders in Six Sigma mini-projects in their own respective areas of expertise. Unlike Black Belts, Green Belts work only part-time on their projects as they have functional responsibilities in their own area of work.

The *Finance Representatives* assist in identifying a project's financial metrics and potential impact, advising the Champion on the approval of projected savings during the define phase of a project. At completion of the project (the end of the project's control phase), he/she will assist in adjustment of projected financial savings due to changes in underlying assumptions (market demand, cost of improvements, etc.). The Finance Representative will also track the actual financial savings of each project for a defined period (usually one year).

1.4 SIX SIGMA TRAINING

All Six Sigma practicing companies enjoy the benefits described earlier, with financial savings in operating costs as an immediate return. In the long run, the workforce will

transform into one that is objectively driven by data in its quest for solutions as Six Sigma permeates through the ranks and functions and is practiced across the organization. To achieve cultural integration, various forms and levels of Six Sigma training must be developed and executed. In addition to the training of Champions and Black Belts (key roles in Six Sigma), appropriate Six Sigma training must be provided across the ranks – from the executives, through the managers, to the engineers and technicians. Administrative functions (finance, human resources, shipping, purchasing, etc.) and non-manufacturing roles (design and development, sales and marketing, etc.) must also be included in the company's Six Sigma outreach.

Champions training typically involves 3 days of training, with primary focus on the following:

- the Six Sigma methodology and metrics;
- the identification, selection and execution of Six Sigma projects;
- the identification, selection and management of Black Belts.

Black Belt training is stratified by the final four phases of a Six Sigma project – Measure, Analyze, Improve and Control. Each phase comprises 1 week of classroom training in the relevant tools and techniques, followed by 3 weeks of on-the-job training on a selected project. The Black Belt is expected to give a presentation on the progress of his/her individual project at each phase; proficiency in the use of the relevant tools is assessed during such project presentations. Written tests may be conducted at the end of each phase to assess his/her academic understanding.

It is the opinion and experience of the author that it would be a mistake to adopt a common syllabus for Black Belts in a manufacturing arena (engineering, manufacturing, quality, etc.) and for those in a service-oriented environment (human resources, information technology, sales and marketing, shipping, etc.). While both groups of Black Belts will require a systematic approach to the identification and eradication of a problem's root causes, the tools required can differ significantly. Customized training is highly recommended for these two major families of application. By the same token, Six Sigma training for hardware design, software design and service design will require more mathematical models to complement the statistical methods.

In addition to the standard 4 weeks of Black Belt training, *Master Black Belt training* includes the Champions training described above (as the Master Black Belt's role bridges the functions between the Black Belt and his/her Champion) and 2 weeks of advanced statistical training, where the statistical theory behind the Six Sigma tools is discussed in greater detail to prepare him/her as the technical expert in Six Sigma.

To facilitate proliferation and integration of the Six Sigma methodology within an organization, appropriate training must be available for all stakeholders – ranging from management who are the project sponsors or Process Owners, to the front-line employees who will either be the team members or enforcers of the proposed solution(s). Such *Green Belt training* is similar to Black Belt training in terms of syllabus, though discussion of the statistics behind the Six Sigma tools will have less depth. Consequently, training is reduced to 4 days (or less) per phase, inclusive of project presentations.

1.5 SIX SIGMA PROJECTS

While Six Sigma tools tend to rely heavily on the use of statistical methods in the analysis within their projects, Black Belts must be able to integrate their newly acquired knowledge with their previous professional and operational experience. Six Sigma may be perceived as fulfilment of the Shewhart–Deming vision:

> *The long-range contribution of statistics depends not so much upon getting a lot of highly trained statisticians into industry as it does in creating a statistically minded generation of physicists, chemists, engineers, and others who will in any way have a hand in developing and directing the production processes of tomorrow.*[5]

The following project is an example of such belief and practice. It demonstrates the deployment of the Six Sigma methodology by a printed circuit board assembly (PCBA) supplier to reduce defect rates to best-in-class levels, and to improve cycle times not only for the pick-and-place process of its surface mount components but also for electrical and/or functional testing. Integration of the various engineering disciplines and statistical methods led to reduction in both direct and indirect material costs, and the design and development of new test methods. Working along with its supply chain management, inventory holding costs were reduced significantly.

1.5.1 Define

In this project, a Black Belt was assigned to reduce the cycle time for the electrical/functional testing of a PCBA, both in terms of its mean and variance. Successful realization of the project would lead to shorter manufacturing cycle time, thus improving the company's ability to respond to customer demands (both internal and external) in timely fashion, as well as offering the added benefit of reduced hardware requirements for volume ramp due to increasing market demand (i.e. capital avoidance).

1.5.2 Measure

To determine the goal for this project, 25 randomly selected PCBAs were tested by six randomly selected testers (Figure 1.6). The average test time per PCBA across all six testers t_{Ave} (*baseline*) was computed, and the average test time per unit for the 'best' tester t_{Best} was used as the *entitlement*. The opportunity for improvement ($\Delta = t_{\text{Ave}} - t_{\text{Best}}$) was then determined. The goal t_{Goal} was then set at 70% reduction of this opportunity, $t_{\text{Goal}} = t_{\text{Ave}} - 0.7\Delta$.

The functional testing of a PCBA comprises three major process steps:

- loading of the PCBA from the input stage to the test bed;
- actual functional testing of the PCBA on the test bed;
- unloading of the tested PCBA to the output stage.

To identify the major contributors of the 'hidden factory' of high mean and variance, 20 randomly selected PCBAs were tested by two randomly selected testers, with each unit being tested three times per tester. The handling time (loading and unloading) and test time (actual functional testing) for each of these tests were measured (see Figures 1.7 and 1.8).

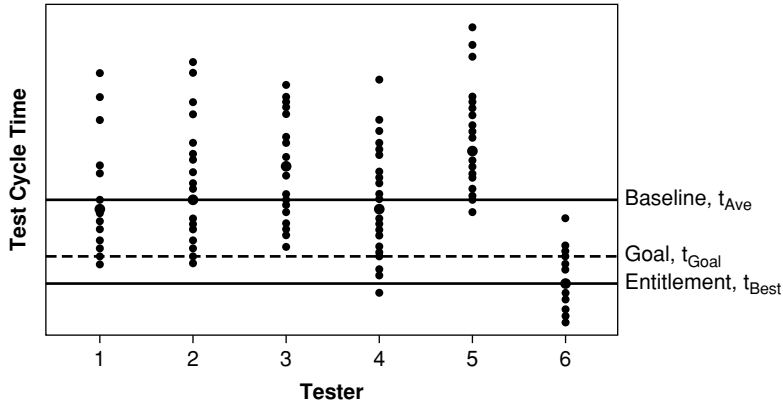


Figure 1.6 Test cycle times for different testers.

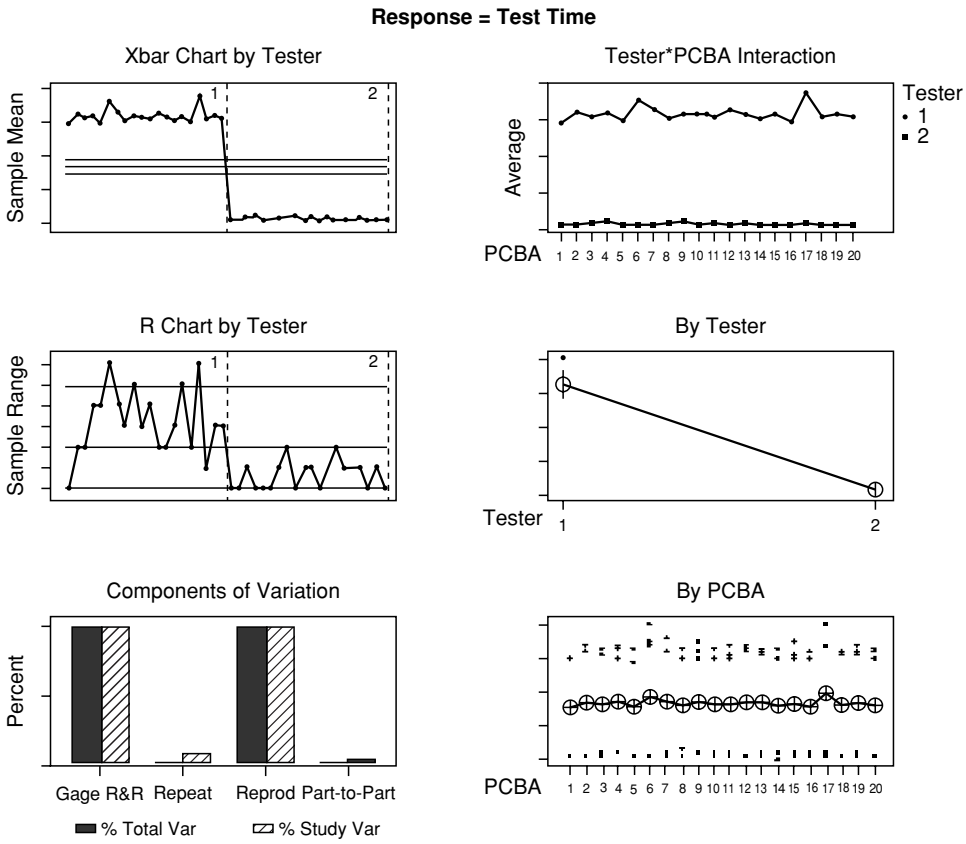


Figure 1.7 Sixpack analysis of test time.

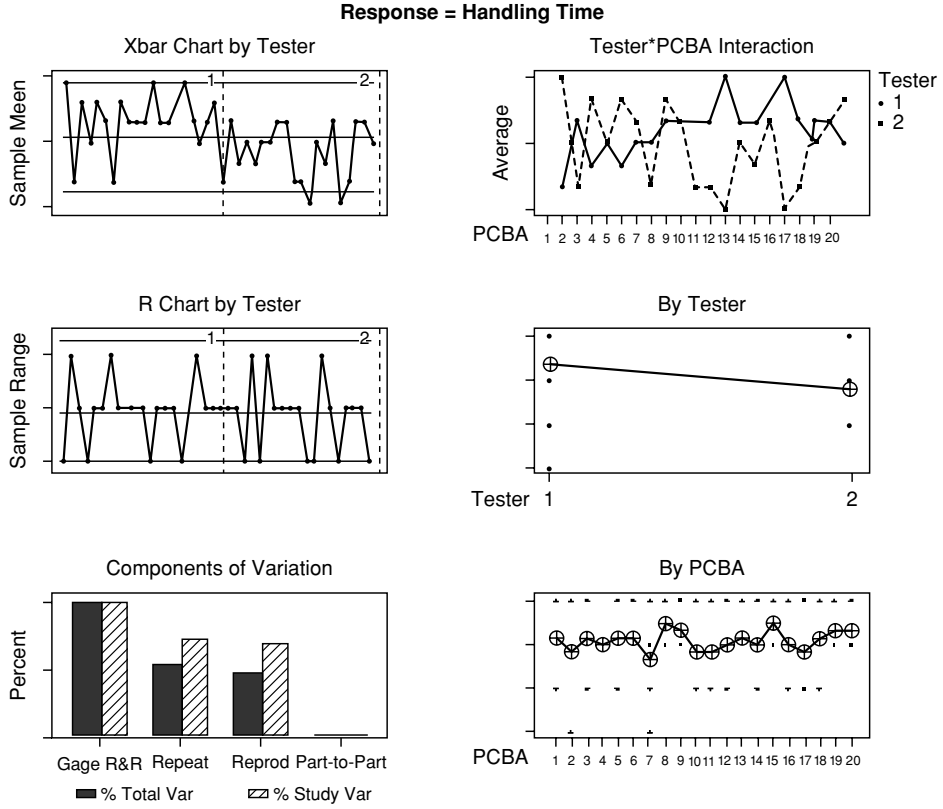


Figure 1.8 Sixpack analysis of handling time.

The following observations were noted:

- Test time was about 6–8 times as large as handling time.
- Variance in handling time between the two testers was negligible.
- Variance in test time between the two testers was significantly different.
- The average test time for Tester 1 was about 25% higher than Tester 2.
- The variance in test time for Tester 1 was nearly 20 times higher than for Tester 2.

The team unanimously agreed to focus their efforts on understanding the causes of variation in test time.

The *fishbone diagram* (also called an *Ishikawa diagram*) remains a useful tool for brainstorming of the various possible causes leading to an effect of concern (Figure 1.9).

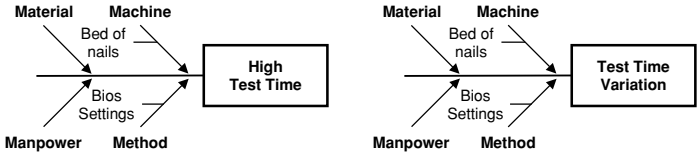


Figure 1.9 Cause-and-effect diagrams for long test time and large variation.

Table 1.1 Cause-and-effect matrix.

Key process input variables	Key process output variables		Score	Rank
	High test time	Test–time variation		
	w_1	w_2		
BIOS settings	r_{11}	r_{12}	S_1	R_1
⋮	⋮	⋮	⋮	⋮
Bed-of-nails	r_{k1}	r_{k2}	S_k	R_k

However, one of its drawbacks is that generally too many possible causes will be listed. To facilitate a somewhat objective selection of important causes for further investigation, the *cause-and-effect matrix* was employed (Table 1.1). A derivative of the *House of Quality*, the importance of the key process output variables – high mean and variance in test time – were reflected in the different weights assigned to them.

The measure s_{ij} reflects the relationship between a *key process input variable* i and the *key process output variable* j. The score of each input variable, $S_i = r_{i1}w_1 + r_{i2}w_2$, was computed and ranked in descending order (i.e. highest score first), with further statistical analysis to be performed on the shortlisted input variables, selected via a *Pareto chart*.

At the end of this phase, the team were confident that they had the solution to their challenge, but they were surprised by what they were to learn.

1.5.3 Analyze

During this phase, statistical experiments and analyses were performed to verify the significance of the shortlisted input variables (Figure 1.10).

Input variables may fall under either of two categories:

- Control factors. Optimum levels for such factors may be identified and set for the purpose, of improving a process's response (e.g. clock speed, BIOS settings).
- Noise factors. Such factors are either uncontrollable, or are costly to control at desired levels (e.g. tester variation).

Regression analysis was performed to identify the effect of clock speed on the PCBA test time (Figure 1.11). While the test time decreased at higher clock speed, there is

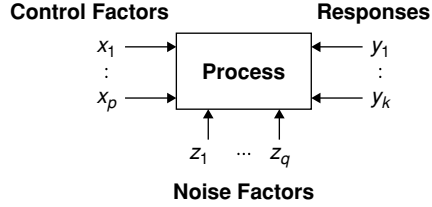


Figure 1.10 Model to facilitate statistical analysis.

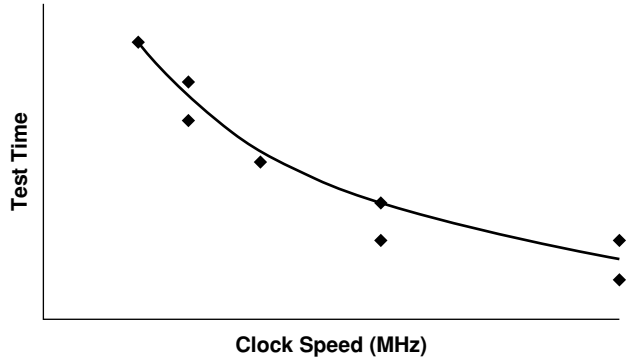


Figure 1.11 Nonlinear relationship between test time and clock speed.

an optimal speed for the existing tester design, beyond which value would not be returned for investment in higher clock speed.

Given the results from the Measure phase, which showed that the variation between testers was highly significant, the team went on to explore two primary sub-systems within a tester, namely the interface card and the test fixture. Five interface cards and six test fixtures were randomly selected for the next experiment; this was to yield results which came as a pleasant surprise.

Before the experiment, it was believed (from experience) that test fixture would result in greater inconsistency due to variation in the contact between the test pins and the test pads, as well as noise due to inductance in the conductors. However, reviewing the results using a *two-way ANOVA Type-II model* reveal that the interface card was the primary cause of variation, not the text fixture.

The *multi-vari chart* in Figure 1.12 illustrates that interface cards A and D can provide robustness against the different test fixtures used, while yielding a shorter test time.

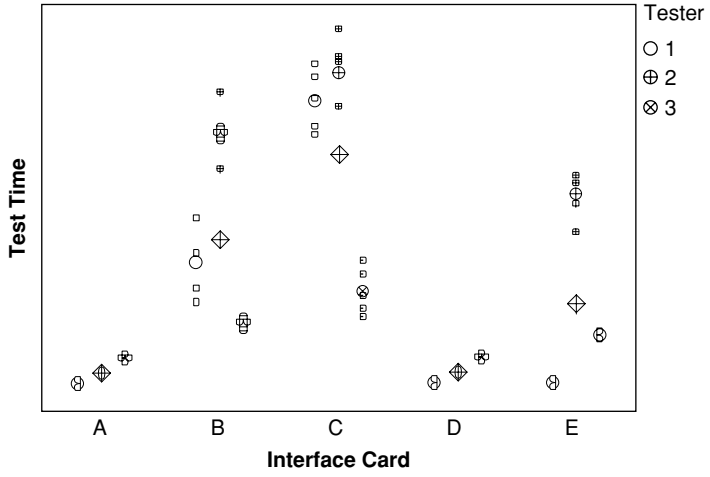


Figure 1.12 Multi-vari chart for test time with different testers and interface cards.

Applying their engineering knowledge, the team narrowed the cause down to the transceiver chip on the card. Examination revealed that cards A and D had transceivers from one supplier, with cards B and C sharing a second transceiver supplier, while card E had its transceiver from a third supplier. Cross-swapping of the transceiver with the interface cards confirmed that the difference was due to the transceiver chip.

1.5.4 Improve

During this phase, the effect of four control factors and one noise factor on two responses was studied.

Response (Y)	y_1	: Average Test Time
	y_2	: Standard Deviation in Test Time
Control Factors (X)	x_1	: Internal Cache
	x_2	: External Cache
	x_3	: CPU Clock Speed
	x_4	: Product Model
Noise Factor (Z)	z_1	: Transceiver on Interface Card

A 2^4 *full factorial design*, with two replicates, blocked by the two transceivers, was employed.

While an optimal combination of control factor levels was identified to minimize both the mean and variance in test time, the results showed that the noise factor (transceiver type) was the largest contributor to improvement.

Engineering analysis was employed to understand the difference between the transceiver chips. Oscilloscope analysis revealed that the 'better' transceiver (from Supplier 1) had a longer propagation delay, that is, it was actually slower than the chip from Supplier 2 (Figure 1.13).

The team verified their finding by acquiring slower transceivers from Supplier 2 (with propagation delay similar to that of Supplier 1). The test time for transceivers

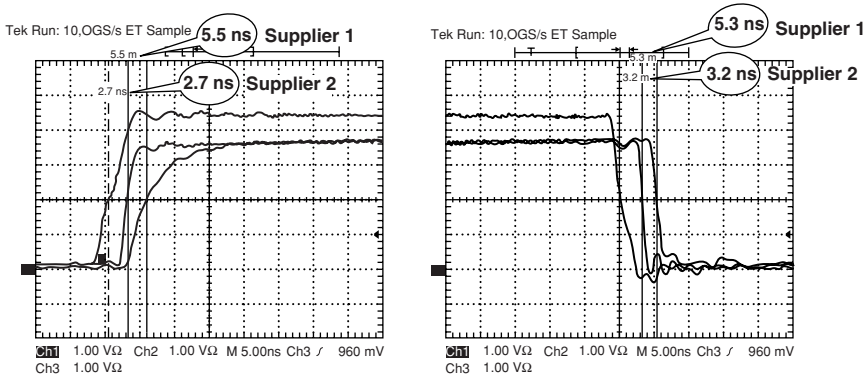


Figure 1.13 Results of oscilloscope analysis on propagation delay for Suppliers 1 and 2.

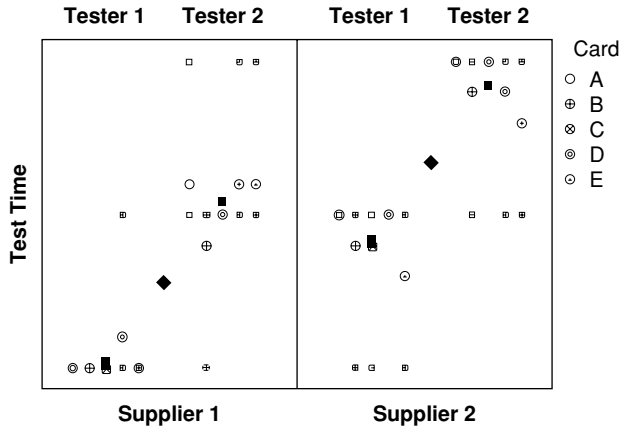


Figure 1.14 Multi-vari chart for test time by testers, interface cards and suppliers.

from both suppliers yielded similar results; verification was performed across two testers and five interface cards (Figure 1.14).

1.5.5 Control

The findings and recommendations were presented to the Process Owner, along with agreed trigger controls. These were documented in a *failure mode and effects analysis* document and *control plan*.

1.5.6 Realize

The results were astounding. Not only did the team exceed the established goal, they actually beat the original entitlement (Figure 1.15). In terms of variation, the variance in test time was reduced to a mere 2.5% of its original value. An unexpected benefit,

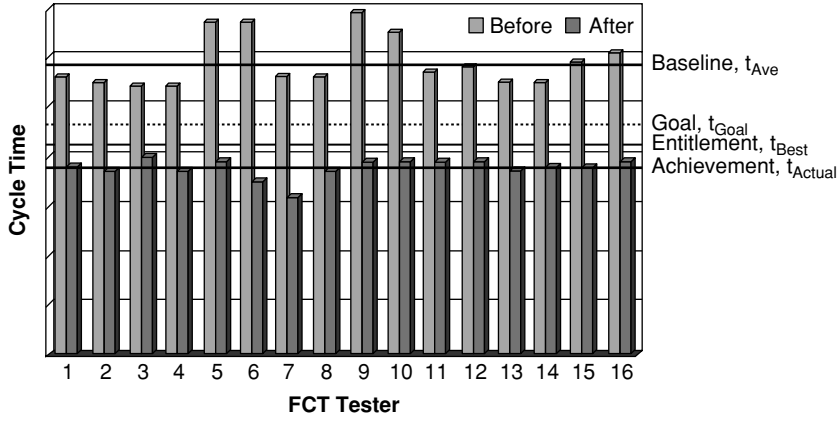


Figure 1.15 A before-and-after comparison of the cycle time.

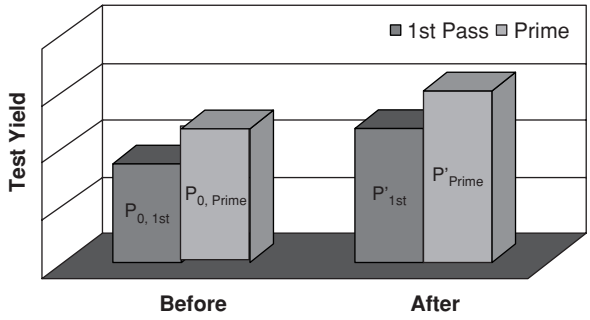


Figure 1.16 A before-and-after comparison of the test yield.

shown in Figure 1.16 was the improvement in first pass yield and prime yield (retest without rework).

The power of Six Sigma was evident in this project. Through the integration of engineering experience/knowledge with the Six Sigma tools and methodology, a data-driven and optimized solution was derived for a highly leveraged problem.

1.6 CONCLUSION

To summarize, Six Sigma involves selecting a highly leveraged problem and identifying the best people to work on it, providing them the training, tools and resources needed to fix the problem, while ensuring them uninterrupted time, so that a data-driven and well-thought-out solution may be achieved for long-term sustenance of profitability.

REFERENCES

1. Harry, M. and Schroeder, R. (2000) *Six Sigma, The Breakthrough Management Strategy Revolutionizing The World's Top Corporations.* New York: Doubleday.
2. Pande, P.S., Neuman, R.P. and Cavanagh, R.R (2000) *The Six Sigma Way*. New York: McGraw-Hill.
3. Ehrlich, B.H. (2002) *Transactional Six Sigma and Lean Servicing*. Boca Raton, FL: CRC Press.
4. Snee, R.D. (1999) Why should statisticians pay attention to Six Sigma? An examination for their Role in the Six Sigma methodology. *Quality Progress*, 32(9), 100–103.
5. Shewhart, W.A. with Deming, W.E. (1939) *Statistical Method from the viewpoint of quality Control.* Washington, DC: The Graduate School, Department of Agriculture.

2

A Strategic Assessment of Six Sigma

T. N. Goh

Six Sigma as a quality improvement framework has been gaining considerable attention in recent years. The hyperbole that often accompanies the presentation and adoption of Six Sigma in industry can lead to unrealistic expectations as to what Six Sigma is truly capable of achieving. In this chapter, some strategic perspectives on the subject are presented, highlighting the potential and possible limitations of Six Sigma applications particularly in a knowledge-based environment. Without delving into the mechanics of the subject in detail, the points raised could be useful to those deliberating on the appropriateness of Six Sigma to their respective organizations.

2.1 INTRODUCTION

Six Sigma as a systematic framework for quality improvement and business excellence has been popular for more than a decade. With its high-profile adoption by companies such as General Electric in the mid-1990s, Six Sigma spread like wild fire in the following years. Detailed accounts of the concepts and evolution of Six Sigma have appeared in several recent issues of *Quality Progress*[1–3] and *Quality Engineering*.[4,5] More recently, some comprehensive discussions on the training of Six Sigma professionals have also been carried in the *Journal of Quality Technology*.[6] Books (in English) on Six Sigma multiplied rapidly, from a handful before 1999,[7–10] about four in 1999,[11–14] to about a dozen in 2000,[15–25] and even more after that,[26–46] not counting the myriad of variations in the form of training kits, instructor's manuals, audio and visual tapes and CDs. The exponential growth of the number of Six Sigma titles is depicted in Figure 2.1.

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

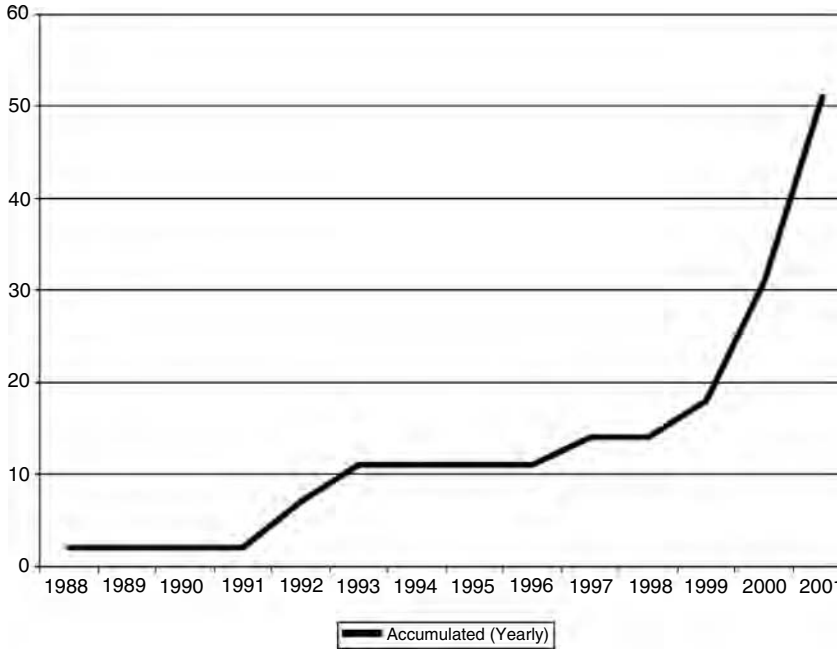


Figure 2.1 Growth of books (in English) on Six Sigma.

Apart from information in the print and audio-visual media, countless Web pages also carry outlines, articles, forums and newsletters on Six Sigma. Six Sigma consulting organizations have mushroomed, each advertising its own, albeit similar, version of Six Sigma. In the face of what might be called the 'Six Sigma phenomenon', a balanced perspective on the subject would be useful before a person or organization takes a decision on whether 'to Six Sigma, or not to Six Sigma' – that is, decides whether to commit financial resources to the adoption of Six Sigma, and on what problems should Six Sigma tools be used.

Six Sigma is unlikely to be a panacea for all quality ills; on the other hand, it must possess sufficient merits for the Six Sigma phenomenon to take hold. In what follows, the essential features of Six Sigma will be highlighted, followed by broad overviews of the potential and limitations of Six Sigma. In particular, the relevance of Six Sigma to a knowledge-based environment is discussed. Technical details of the subject will not be elaborated as they are commonly available; emphasis will be placed instead on strategic considerations.

2.2 SIX SIGMA FRAMEWORK

The practice of Six Sigma takes the form of projects conducted in phases generally referred to recognized as define–measure–analyze–improve–control (DMAIC). After the define phase of a project, key process characteristics are identified, studied and benchmarked in the measure and analyze phases. Then in the improve phase a process

is changed to give a better or optimized performance. Finally, the control phase then ensures that the resulting gains are sustained beyond the completion of the project.

The use of statistical thinking[47] is a common thread through these phases, with measured data providing an indispensable proxy for realities and facts. Thus, Harry and Schroeder presented Six Sigma as 'a disciplined method of using extremely rigorous data gathering and statistical analysis to pinpoint sources of errors and ways of eliminating them'.[13] The levels of competence of personnel executing the projects are classified or labeled as 'Master Black Belts', 'Black Belts', 'Green Belts' and so on. Further Six Sigma training and implementation details are described in various references.[4–6]

2.3 SIX SIGMA FEATURES

There are several features that distinguish Six Sigma from other quality improvement initiatives. First is the DMAIC framework, where techniques such as quality function deployment, failure mode and effects analysis, design of experiments and statistical process control (SPC) are integrated into a logical flow. Gone are the days when these techniques learned and used in a disjointed manner or disconnected sequence.

The second feature is the approach advocated for Six Sigma. Implementation is supposed to be 'top-down' from the CEO, rather than something promulgated by the quality assurance people, human resources department, or at ground level by a quality control circle. The experience of General Electric best exemplifies this approach.

Although Six Sigma has a substantial number of statistical techniques that are traditionally used in the manufacturing industry, its application is not limited to operations in manufacturing. The use of Six Sigma in transactional or commercial situations is actively promoted, rendering a new dimension to service sector quality in terms of rigor of problem solving and performance improvement.

Associated with the wider scope of application is customer focus. This is emphasized repeatedly in Six Sigma in terms of issues that are *critical to quality* (CTQ); improvements will make sense only if they are directly related to some CTQs. Thus, in contrast to some of the inward-looking efforts of ISO or QS certification, Six Sigma is much more sensitive to requirements for customer satisfaction.

In terms of organization, Six Sigma stresses the project-by-project feature of its implementation; this is distinct from the valid but nebulous 'quality is free' concept or 'company-wide quality improvement' efforts in the past. A project has a concrete objective, a beginning and an end, and provides opportunities for planning, review and learning. Indeed, projects are featured prominently in formal Six Sigma training programs, something not often seen in quality-related training activities in the past.

The outcomes of Six Sigma projects are usually required to be expressed in financial terms. This leads to a direct measure of achievement which most people understand – not just the project members. Compared to exhortations to achieve zero defects or to do it right the first time, in which the outcome has to be strictly black or white (success or failure), financial bottom lines provide a much better measure of the impact of improvements as well as a vivid calibration of progress.

Another important feature of Six Sigma is the elaborate training and certification processes that result in Black Belts, Green Belts, and so on. This is in contrast to

the *ad hoc* or one-off nature of on-the-job training in the past, where there was no commonly recognizable way of designating an individual's competence or experience in effecting quality improvement.

All these features – framework, approach, application, focus, organization, results and personnel – are important contributors to the effectiveness of Six Sigma. Coupled with project management techniques, together they provide a comprehensive framework for effective application of statistical thinking and methodologies for problem solving and demonstrable measures of improvement.

2.4 SIX SIGMA: CONTRASTS AND POTENTIAL

As pointed out, statistical thinking and statistical methodologies constitute the backbone of Six Sigma. MINITAB, a commonly used software package, describes Six Sigma as 'an information-driven methodology for reducing waste, increasing customer satisfaction and improving processes, with a focus on financially measurable results'.[48] The idea of information-based improvement has now been extended to design activities, in the form of design for Six Sigma (DFSS).[46,49] DFSS – typically in the form of identify–design–optimize–validate (IDOV) – aims to design products, services and processes that are 'Six Sigma capable', emphasizing the early application of Six Sigma tools and the fact that as far as defect elimination goes, prevention is better than cure. The results are a far cry from the days when quality had to depend on testing and inspection (T&I), or perhaps SPC.

It can be seen that the emphasis of quality improvement has been moving gradually upstream over the years: from T&I on the product to SPC on the process, then Six Sigma on the system, and finally DFSS as a pre-emptive move for achieving the desired performance. Without delving into the details, which are available elsewhere,[49–52] the differences among these approaches are summarized in Table 2.1. Certainly Six Sigma and DFSS represent a far more fundamental approach to problem solving and problem anticipation, respectively, in any given situation.

There are many factors that contribute to the potential of Six Sigma, of which the critical ones are as follows:

1. top-down initiation of a serious quality journey (not a book-keeping exercise);
2. hierarchy of expertise and execution (Champions, Black Belts, etc.);
3. structured deployment of tools (DMAIC);
4. customer focus (in contrast to inward-looking standardization);
5. clear performance metric (sigma levels; defects per million opportunities (DPMO));
6. fact-based decisions (not procedure- or judgment-based);
7. application of statistics (analytical, not will power);
8. service as well as engineering applications (thus extending the horizon of statistical thinking);
9. recognized time effects in process analysis (with explicit provisions for short-term and long-term variations);
10. result-oriented (project by project; project duration of 3–6 months makes progress tangible);

Table 2.1 Progress in efforts for performance improvement.

	Methodology	T&I	SPC	Six Sigma	DFSS
1.	Approach	Defect detection	Defect prevention	Defect avoidance	Value creation
2.	Method	Sampling plans	Control charts	DMAIC	IDOV
3.	Focus	Product	Process	Project	System
4.	Information	Static	Dynamic	Varied	Uncertainties
5.	Medium	Observation	Data	Knowledge	Perspectives
6.	Nature	Passive	Defensive	Active	Pre-emptive
7.	Deployment	Exit point	Downstream	Midstream	Upstream
8.	Application	Isolated	On-line	Off-line	Organization-wide
9.	Format	As needed	Continuous	Project by project	Subject by subject
10.	Operation	Single location	Single function	Cross-function	Business-wide
11.	Execution	Prescriptive	Rule-based	Needs-driven	Proactive
12.	Criterion	Conformance	Stability	Optimality	Predictability
13.	Improvement	Irrelevant	Absent	Incremental	Fundamental
14.	Problems	Unsolved	Contained	Understood	Anticipated
15.	Solutions	Unavailable	*Ad hoc*	Remedial	Built-in
16.	Result	Damage control	Capability	Sigma level	Robustness
17.	Framework	Instantaneous	Short term	Long term	Life cycle
18.	Customer reaction	Acceptance	Satisfaction	Appreciation	Trust
19.	Gains	None	Confidence	Savings	Profit
20.	Enhancement	Production	Engineering	Bottom line	Market share
21.	Requirements	Unsophisticated	Procedural	Organizational	Cultural
22.	Core skills	Procedures	Analysis	Communication	Synthesis
23.	Leaders	Technicians	Engineers	Managers	Chief executives
24.	Applicability	Traditional	Modern	Contemporary	Current
25.	Popularity started	1940s	1970s	1990s	2000s

Commonality: management of variabilities with statistical thinking

11. business-oriented (achievements often required to be expressed in financial terms);
12. good timing (coming at a time when personal computing hardware and statistical software packages have become widely available, making pervasive implementation possible).

2.5 SIX SIGMA: INHERENT LIMITATIONS

Six Sigma was first applied to industrial manufacturing processes in which defects can be clearly defined as well as measured; the extent of improvement likewise has to be quantifiable. It is well known that Six Sigma derives its capability for process

improvement largely from statistical tools that relate input–output data and make use of analytical transfer functions for optimization studies. When extended to non-manufacturing processes, the premise is again that the output is wanted by the customer and desired to be uniform around some specified target. Regardless of the application, the use of quantitative measures is stressed throughout: one adverse consequence of requiring successes to be judged by numbers is, as Galbraith pointed out, that 'to many it will always seem better to have measurable progress toward the wrong goals than unmeasurable progress toward the right ones'.[53]

With Six Sigma starting out as a defect prevention and error avoidance scheme, one still finds today, in promotions for its adoption, arguments such as if airlines do not operate at Six Sigma level, there will be so many crashes a month; if power companies are not at Six Sigma level, there will be so many hours of blackouts per week; if utilities are not at Six Sigma level, there will be so much unsanitary drinking water per day, and so on. Such 'illustrations', if not emotional blackmail, could simply be a reflection of naïveté in problem solving. In reality, when performance is expressed in terms of the common measure of DPMO, it would be unwise to assume that all non-defects are equally good or even desirable, and that all defects are equally damaging – for example, unsold electricity is profit lost to the power supplier, a poor aircraft landing could take many forms, and a defective hospital procedure could result in anything from a slight annoyance to a life-threatening situation.

As Six Sigma is increasingly being touted as the route to organizational and business excellence, it must be noted that one would be grossly off the mark if conformance to numerical specifications and minimization of errors were to be forced upon a thriving, forward-looking enterprise such as one engaged in research and development. If no mistake is to be made, the first step would have to be suppression of all innovative thinking and exploratory activities among the staff. Six Sigma can serve as a prescription for conformance – safe landing of an airplane, uninterrupted electricity supply, sanitary drinking water, successful operation in a hospital, etc. – but hardly a formula for creativity, breakthrough or entrepreneurship.

There is yet another aspect of Six Sigma as it is known and practiced today that calls for attention. Partly owing to the pressure to show results and demonstrate successes, many Six Sigma practitioners tend to work on problems that are related to the 'here' and 'now' around them. There is no guarantee that the problem solving or process optimization efforts led by Champions and Black Belts are, from larger perspectives, well conceived or well placed. To make the point, consider the string quartet on board the *Titanic*: at some stage it might be brilliant at technical delivery, error avoidance, team work and customer satisfaction – nevertheless it was doomed right where it was doing all these. In other words, a larger picture or time frame could show whether a Six Sigma initiative is meaningful or worthwhile.

In the increasingly globalized environment today, new products and services are constantly needed in anticipation of customer requirements, cultural trends, changing lifestyles, new technologies or unexpected business situations. Customization – that is, i.e. variety – is of increasing importance relative to uniformity and predictability. Unfortunately it is not uncommon to see Black Belt projects formulated with an internal focus or dominated by local concerns and prevalent measures of performance.

Naturally, not all of the above aspects are necessarily material drawbacks in any given Six Sigma journey, but they do provide a reminder that Six Sigma cannot be a

universal solution for any organization in any situation. The list below summarizes some common attributes in a typical Six Sigma implementation; in fact, they could well serve as pointers for better practice:

1. It relies on the measurable (with a tendency to avoid the unquantifiable in project selection).
2. Attention is paid to repetitive output (with lack of methodology for innovative or irregular outcomes).
3. It is focused on error prevention (not gains from creativity or imagination).
4. It is founded upon unrealistic mathematical statistics (such as the normal distribution and 1.5 sigma shifts).
5. It is mostly concerned with basic CTQ (i.e. lack of attention to unexpected or 'delighting' CTQ as in the Kano quality model).[54]
6. It studies only current, static CTQ (with little reference to varied customer expectations or lifestyles; it does not anticipate technological, social, or business changes).
7. It is usually based on one CTQ (i.e. single rather than multiple or balanced CTQ in a given project).
8. There is practically no emphasis on self-learning or future knowledge acquisition in personnel training.
9. It is unsuitable for creative or interpretive work (e.g. architectural design, artistic performance).
10. It is not a means to promote intellect, creativity, passion, enterprise or self-renewal.
11. It emphasizes the priorities of the organization (rather than the growth of people, e.g. talent development or continuous learning on the part of Black Belts and Green Belts; personnel are mechanically classified in terms of terminal qualifications).
12. It tends to be preoccupied with internal objectives (with no reference to social mission or responsibility).

In this light, less conventional views of Six Sigma have appeared in various sources expressing, for example, skepticism,[55] alternative interpretations,[56] or suggestions of its possible future.[57] To take the issue further, an examination will be made next of environments in which Six Sigma could be found inappropriate, infeasible, or simply irrelevant.

2.6 SIX SIGMA IN THE KNOWLEDGE ECONOMY

The greatest obstacle faced by Six Sigma practitioners is the predominance of the overarching philosophy of defect prevention. This is especially true in situations where, as described elsewhere,[58] knowledge is being acquired, created, packaged, applied and disseminated. In a knowledge-based organization and, by extension, a knowledge-based economy, the culture tends to be shaped by the following:

1. Knowledge work is characterized by variety, exception, novelty and even uncertainty, rather than regularity and predictability.
2. Productivity and valued added, rather than degree of conformance, constitute the objectives as well as challenges in knowledge management.

3. The processes of knowledge management and knowledge worker involvement are significantly different from those in repetitive operations and administration, the targeted areas for Six Sigma applications.
4. Knowledge workers, such as research scientists, urban designers and music composers, tend to resist or detest structured approaches, and enjoy the freedom of exploring the unknown.
5. Knowledge workers conceptualize, design, execute and improve their work with their own experience and judgment; spontaneity and autonomy could add to their performance.
6. Many knowledge workers are motivated by the process more than the outcome, and could even be rewarded on the basis of their input rather than output.
7. It is impractical to set targets such as '*X* times better' or calibrate improvements in terms such as sigma levels in knowledge management; large portions of the achievements of knowledge workers could be intangible in nature.
8. As an economy develops, accompanied by changes that come with technology and globalization, there is impetus for migration of human resources from the operational and administrative arena – fertile grounds for DMAIC – to knowledge-based activities and organizations.
9. Knowledge workers and knowledge societies sustain themselves through continuous learning and self-renewal, with constant changes in processes and process outputs.
10. Ultimately, it is the acquisition, creation and application of knowledge that would fuel the competitiveness of a knowledge organization or society.

The rise of the knowledge worker and the emergence of the knowledge society, as well as the consequent impact on the delivery and assessment of performance, have been well discussed in the context of management;[59] it was pointed out that formal education and theoretical knowledge, not adherence to standard operating procedures or apprenticeship, are prerequisites for an effective knowledge worker. Expressing the level of performance and its improvement via some definitive metric would fail to make sense in a knowledge-based organization such as a university. Thus, for example, one would be hard put to calculate a meaningful sigma score in each of the following cases:

- ratings from student feedback on the content and delivery of a course;
- graduation or drop-out rate of a given program of study;
- average time taken by fresh graduates to get suitable employment;
- number and nature of continuing education programs offered as a community service;
- sizes, facilities, nature and range of services of libraries;
- proportion of submitted research papers accepted for publication in journals;
- number of patents filed out of *X* million dollars of research funding;
- Number of successful industrial collaboration projects in an academic department;
- annual turnover of administrative staff in a university;
- proportion of tenured faculty in a particular college;
- service rendered by faculty to professional organizations;
- reputation of the university as perceived by international peers.

In fact, administratively, conformance is required in Six Sigma not just with respect to process output but also in terms of employee participation. Regimentation via a hierarchy of Master Black Belts, Black Belts, and so on would go against the very culture in a community of academics and researchers, even students. The bottom-line approach taken in Six Sigma cannot be the driving factor in the pursuit of academic excellence. Cynicism, well before anything else, would be the first reaction if the leadership of a university were to tell everyone 'Take Six Sigma – or this organization is not for you', a proclamation made famous by Jack Welch as CEO of General Electric. It is little wonder that, to date, no university has proclaimed itself a Six Sigma institution.

2.7 SIX SIGMA: IMPROVING THE PARADIGM

Six Sigma has been widely publicized in recent years as the most effective means to combat quality problems and win customer satisfaction. As a management initiative, Six Sigma is best suited to organizations with repetitive operations for specified outcomes. It aims at preventing non-conformance to defined formats and contents of outcomes, generally identified as defects in products and errors in transactions. The success of Six Sigma applications is reported in terms of sigma levels and the benefits reflected by some financial bottom line. A number of high-profile companies have attributed to Six Sigma the substantial gains in the few years over the turn of the century – coincidentally also a high-growth period for the broad US economy.

Several facets of Six Sigma have been analyzed in the previous sections, with highlights on success factors and possible limitations. In the face of overwhelming voices on the merits and power of Six Sigma, the fact remains that Six Sigma is most effective when an organization already has a firm idea of what forms of products and services are in alignment with the organization's goals and customer expectations. Six Sigma is suited to problems in which the output can be readily measured. The methodology is meant to be implemented by a hierarchy of specially trained personnel – the 'Belts' of various colors. Thus there is now a fast-growing industry of Six Sigma consultants and training programs, with widespread certification activities conducted in an unregulated variety of ways.

It could be argued that Six Sigma is relevant when consistency of performance is valued and its maintenance desirable. Six Sigma is called for when avoidance of non-conformance is of higher priority than breakthrough and creativity. Thus, while Six Sigma has its place in securing predictable product and service characteristics in businesses, its very nature would run counter to the culture of creativity and innovation in any vibrant, innovation-oriented enterprise. Six Sigma is commonly applied to address what has gone wrong, but not what is beyond the current perception of what is CTQ. Nor does the Six Sigma framework explicitly deal with the worth of knowledge, imagination, innovation, passion or dedication. It is all too easy to avoid mistakes or failures by not trying anything novel: when obsessed with error avoidance, one's attention and energy would tend to be diverted from exploratory pursuits and endeavors for value creation.

It may be said, in conclusion, that instead of unreserved raving about 'all things Six Sigma', a balanced view of the strengths and weaknesses of Six Sigma is in order. It is important to be aware of the nature of situations in which Six Sigma could or

should be applied. As an approach to organizational excellence, Six Sigma, as it has been practiced, can certainly be enhanced. For example, methodologies of systems engineering[60] can be fruitfully used to break away from the narrow attention span of the 'here and now'. The systems perspective is effective in handling multiple CTQs so that, with a macro-view, suboptimization due to rigid concerns in specific projects can be avoided. Scenario-based planning is another useful approach for coping with dynamic market demands, latent needs, transformed living conditions, varied cultures and changing CTQs. Indeed, a CTQ valid today is not necessarily a meaningful one tomorrow; shifting social, economic and political scenes would make it imperative that except for immediate, localized projects, all CTQs should be critically examined at all times and refined as necessary.

In a mature organization, employees grow with organizational successes. In people-centered operations, in particular, staff development should go hand-in-hand with prevailing Six Sigma applications. This will help reap the benefits of continuous learning, knowledge accumulation, group innovation and creativity. At the collective level, conscious efforts could also be made to meet the requirements of the organization's social mission and obligations. The synergy of such endeavors and Six Sigma based activities and programs could lead to results larger than the sum of isolated or unconnected efforts, and help ensure maintenance or improvement of overall organizational performance in uncertain times.

As the scope of Six Sigma widens in the years to come, it is highly probable that new and more comprehensive variants of the framework and methodologies will gradually emerge, overcoming the possible shortcomings raised here. It is particularly important that suitable adaptations be developed so that Six Sigma can make itself relevant in the face of the very nature of a knowledge enterprise or knowledge-based economy. That indeed would be in line with the 'change' element in Six Sigma culture and an exemplification of *continuous improvement* in the most useful way, an outcome all stakeholders of Six Sigma would be happy to see.

REFERENCES

1. Harry, M.J. (1998) Six Sigma: A breakthrough strategy for profitability. *Quality Progress*, 31 (5), 60–64.
2. Hoerl, R.W. (1998) Six Sigma and the future of the quality profession. *Quality Progress*, 32(6), 35–42.
3. Snee, R.D. (2000) Six Sigma improves both statistical training and processes. *Quality Progress*, 33(10), 68–72.
4. Hahn, G.J., Doganaksoy, N. and Hoerl, R.W. (2000) The evolution of Six Sigma. *Quality Engineering*, 12(3), 317–326.
5. Snee, R.D. (2000) Impact of Six Sigma on quality engineering. *Quality Engineering,* 12(3), ix–xiv.
6. Hoerl, R.W. (2001) Six Sigma black belts: What do they need to know? (with discussion). *Journal of Quality Technology*; 33(4), 391–435.
7. Harry, M.J. (1988) *The Nature of Six Sigma Quality*. Schaumburg, IL: Motorola University Press.
8. Harry, M.J. and Lason J.R. (1992) *Six Sigma Producibility Analysis and Process Characterization*. Reading, MA: Addison-Wesley.
9. Prins, J. (1993) *Six Sigma Metrics.* Reading, MA: Addison-Wesley.
10. Harry, M.J. (1997) *The Vision of Six Sigma*, Vols 1–8. Scottsdale, AZ: Sigma Publishing.

11. Breyfogle III, F.W. (1999) *Implementing Six Sigma: Smarter Solutions Using Statistical Methods*. New York: John Wiley & Sons, Inc.
12. Pyzdek, T. (1999) *The Complete Guide to Six Sigma*. Milwaukee, WI: Quality Press.
13. Harry, M.J. and Schroeder R. (1999) *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations*. New York: Doubleday.
14. Perez-Wilson, M. (1999) *Six Sigma: Understanding the Concept, Implications and Challenges*. Scottsdale, AZ: Advanced Systems Consultants.
15. Brue, G. (2000) *Six Sigma for Leadership*. Pagosa Springs, CO: Morningstar Communications Group.
16. Brue, G. (2000) *Six Sigma for Team Members*. Pagosa Springs, CO: Morningstar Communications Group.
17. Pande, P.S., Neuman, R.P. and Cavanagh R.R. (2000) *The Six Sigma Way*. New York: McGraw-Hill.
18. Naumann, E. and Hoisington, S.H. (2000) *Customer Centered Six Sigma*. Milwaukee, WI: Quality Press.
19. Eckes, G. (2000) *The Six Sigma Revolution*. New York: John Wiley & Sons, Inc.
20. Star, H. and Snyder, S.J. (2000) *Understanding the Essentials of the Six Sigma Quality Initiative*. Bloomington, IN: 1st Books Library.
21. Breyfogle III, F.W., Cupello, J.M. and Meadows B. (2001) *Managing Six Sigma*. New York: John Wiley & Sons, Inc.
22. Tennant, G. (2001) *Six Sigma: SPC and TQM in Manufacturing and Services*. Aldershot: Gower.
23. Pyzedek, T. (2001) *The Six Sigma Handbook: A Complete Guide for Green Belts, Black Belts, and Managers at All Levels*. New York: McGraw-Hill.
24. Arthur, L.J. and Arthur, J. (2000) *Six Sigma Simplified*. Denver, CO: Lifestar.
25. Oriel Inc. (2000) *Guiding Successful Six Sigma Projects*. Madison, WI: Oriel Inc.
26. Brassard, M. and Ritter, D. (2001) *Sailing through Six Sigma*. Milwaukee, WI: Quality Press.
27. Chowdhury, S. (2001) *The Power of Six Sigma*. Chicago: Dearborn Trade.
28. Eckes, G. (2001) *Making Six Sigma Last*. New York: John Wiley & Sons, Inc.
29. Pande, P.S., Cavanagh, R.R. and Neuman, R.P. (2001) *The Six Sigma Way Team Field Book*. New York: McGraw-Hill.
30. Munro, R.A. (2002) *Six Sigma for Operators*. Milwaukee, WI: Quality Press.
31. Stamatis, D.H. (2001) *Foundations of Excellence, Six Sigma and Beyond*, Vol. I. Delray Beach, FL: Saint Lucie Press.
32. Stamatis, D.H. (2001) *Problem Solving and Basic Statistics, Six Sigma and Beyond*, Vol. II. Delray Beach, FL: Saint Lucie Press.
33. Stamatis, D.H. (2001) *Statistics and Probability, Six Sigma and Beyond*, Vol. III. Delray Beach, FL: Saint Lucie Press.
34. Stamatis, D.H. (2001) *Statistical Process Control, Six Sigma and Beyond*, Vol. IV. Delray Beach, FL: Saint Lucie Press.
35. Pande, P. and Holpp, L. (2001) *What is Six Sigma?* Maidenhead: McGraw-Hill.
36. Bhote, K.R. (2001) *The Ultimate Six Sigma: Beyond Quality Experience*. Maidenhead: AMACOM.
37. Breyfogle III, F.W., Enck, D., Flories, P. and Pearson, T. (2001) *Wisdom on the Green: Smarter Six Sigma Business Solutions*. Austin, TX: Smarter Solutions.
38. Munro, R.A. (2001) *Six Sigma for the Shop Floor*. Milwaukee, WI: ASQ Quality Press.
39. Lowenthal, J.N. (2001) *Six Sigma Project Management*. Milwaukee, WI: ASQ Quality Press.
40. Mills, C., Wheat, B. and Carnell, M. (2001) *Leaning into Six Sigma*. Holt, MI: Publishing Partners.
41. Keller, P.A. (2001) *Six Sigma Deployment*. Tucson, AZ: Quality Publishing.
42. Eckes, G. (2002) *Six Sigma Team Dynamics*. Chichester: John Wiley & Sons, Inc.
43. George, M.L. (2002) *Lean Six Sigma: Combining Six Sigma Quality with Lean Production Speed*. New York: McGraw-Hill.
44. Brue, G. (2002) *Six Sigma for Managers*. New York: McGraw-Hill.
45. Ehrlich, B.H. (2002) *Transactional Six Sigma and Lean Servicing*. Delray Beach, FL: Saint Lucie Press.

46. Tennant, G. (2002) *Design for Six Sigma: Launching New Products and Services without Failure*. Aldershot: Gower.
47. Hoerl, R.W. and Snee, R.D. (2002) *Statistical Thinking Improving Business Performance*. Pacific Grove, CA: Duxbury Press.
48. Six Sigma enigma – taking the mystery out of the methodology. http://www.onesixsigma.com/Sponsors/MINITAB/Articles%20and%20White%20Papers/Six%20Sigma%20Enigma.aspx.
49. Goh, T.N. (2002) The role of statistical design of experiments in Six Sigma: perspectives of a practitioner. *Quality Engineering*, 14, 661–673.
50. Grant, E.L. and Leavenworth, R.S. (1996) *Statistical Quality Control* (7th edn). New York: McGraw-Hill.
51. Box G.E.P., Hunter W.G., Hunter J.S. (1978) *Statistics for Experimenters*. Wiley: New York.
52. Goh, T.N. (1993) Taguchi methods: Some technical, cultural and pedagogical perspectives. *Quality and Reliability Engineering International*, 9, 185–202. See also Chapter 18, this volume.
53. Galbraith, J.K. (1967) *The New Industrial State*. Boston: Houghton Mifflin.
54. Kano, N., Seraku, N., Takahashi, F. and Tsuji, S. (1984) Attractive quality and must-be quality. *Hinshitsu*, 14(2), 39–48.
55. Lee, C. (2001) Why you can safely ignore Six Sigma. *Fortune*, 22 January.
56. Schrage, M. (2001) Make no mistake? *Fortune*, 24 December.
57. Montgomery, D.C. (2001) Beyond Six Sigma. *Quality and Reliability Engineering International*, 17(4), iii–iv.
58. Davenport, T.H., Jarvenpaa, S.L. and Beers, M.C. (1966) Improving knowledge work processes. *Sloan Management Review*, 37(4), 53–65.
59. Drucker, P. (1994) The age of social transformation. *Atlantic Monthly*, 274, 53–80.
60. Sage, A.P. and Armstrong, J.E. (2000) *Introduction to Systems Engineering*. New York: John Wiley & Sons, Inc.

3

Six Sigma SWOT

T. N. Goh and L. C. Tang

Six Sigma as an initiative for quality improvement and business excellence has been a subject of intense study and discussion since the early 1990s. Although concepts and templates such as DMAIC have been frequently presented and advocated, there exist a number of different perspectives as to what Six Sigma is fundamentally capable of. In this chapter we examine the strengths, weaknesses, opportunities, and threats associated with the application of the common version of Six Sigma. Some statistical, engineering, and management aspects are considered. Thus, a realistic technical as well as management review of Six Sigma is made, twenty years after its inception.

3.1 INTRODUCTION

Six Sigma as a quality improvement framework gained the attention of industry over a decade ago. Not only manufacturing companies, but also service organizations such as financial and educational institutions have started to embrace Six Sigma concepts and strategies. Success stories emerging from Motorola, General Electric, Seagate Technologies and Allied Signal have enticed organizations to adopt Six Sigma as the ultimate tool towards performance perfection and customer satisfaction.[1] However, Six Sigma has always been subject to debate and criticism. It is therefore useful for quality practitioners to take an objective view of the pros and cons of Six Sigma as it is commonly practiced, since any serious commitment to it would entail considerable corporate resources and management reorientation. This chapter examines the strengths, weaknesses, opportunities and threats (SWOT) of the Six Sigma strategy, and offers both technical and management viewpoints on the subject for those who are seriously considering embarking on Six Sigma implementation.

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

3.2 OUTLINE OF SIX SIGMA

Six Sigma can be traced back to the 1980s when Motorola, Inc. developed and implemented a new quality program based on the concept of variation management. Originally, it was a way to improve performance to the theoretical level of 3.4 defects per million manufactured units or transactions. The program was designed to identify, measure, reduce, and control the variation found in any realistic environment. Its application then spread from Motorola to other companies such as General Electric, Allied Signal and Seagate Technologies.[2]

In general, Six Sigma deals with the fact that process and product variation is usually a strong factor affecting manufacturing lead times, product and process costs, process yields, product quality, and, ultimately, customer satisfaction. The traditional Six Sigma process improvement framework is based on a rigorous process improvement methodology that has the following stages: define, measure, analyze, improve, and control (DMAIC). These key stages are defined as follows:

Define. Define the problem to be solved, including customer impact and potential benefits.
Measure. Identify the critical-to-quality characteristics (CTQs) of the product or service. Verify measurement capability, designate the current defect rate as baseline, and set goals for improvement.
Analyze. Understand the root causes of defects; identify key process input variables (KPIVs) that cause defects.
Improve. Quantify the influences of the KPIVs on the CTQs, and identify acceptable limits of these variables; modify the process to stay within these limits, thereby reducing defect levels in the CTQs.
Control. Ensure that the modified process now keeps the key process output variables (KPOVs) within acceptable limits, in order to maintain the gains in the long term.

Successful Six Sigma implementation in any organization is a top-down initiative implemented by a hierarchy of trained personnel designated as Champions, Master Black Belts, Black Belts and Green Belts. It is an improvement effort enforced top-down and not expected to be a bottom-up phenomenon, although buy-in at the grass-roots level will contribute to its success.

The implementation of Six Sigma in manufacturing is intended to eliminate almost all defects, rework and scrap. But eventually it should be about more than just delivering products without defects. It should bring processes under statistical control and lead to quality based on design rather than inspecting for defects at the end of the process. Ultimately it should result in maximizing equipment utilization time and optimizimg cycle time.

3.3 SWOT ANALYSIS OF SIX SIGMA

An objective analysis will now be made of the various aspects of Six Sigma. The SWOT format will be used, addressing the strengths, weaknesses, opportunities and threats associated with the process.

3.3.1 Strengths

3.3.1.1 Customer focus

Customer focus is the core of quality and the ultimate goal of any successful process. In a typical Six Sigma program for process improvement, the aim is to build what the customers want, and this is reflected in the product CTQs. Improvements are defined by their impact on customer satisfaction, achieved through the systematic framework and tools of Six Sigma.

3.3.1.2 Data-driven and statistical approach to problem solving

A strong focus on technically sound quantitative approaches is the most important feature of Six Sigma. Six Sigma is firmly rooted in mathematics and statistics. Statistical tools are used systematically to measure, collect, analyze and interpret data and hence identify the directions and areas for process improvement. The once-popular quality program, total quality management (TQM), seemed to be little different from Six Sigma in the view of many quality practitioners who found both systems have much in common.[3] However, Six Sigma adopts a systematic quantitative approach that overcomes the difficulties incurred by the general and abstract guidelines in TQM; these guidelines could hardly be turned into a successful deployment strategy.

3.3.1.3 Top-down support and corporate-wide culture

Six Sigma requires a top-down management approach. The initiative must come from top management and be driven through every level of the organization. It is not simply a matter of top management approving the budget for a Six Sigma implementation and expecting other levels simply to get on with it. In such a situation, the project would be doomed to failure from the start.[4] With this top-down approach, a sense of 'urgency' will be felt by members of Six Sigma projects and their work will be taken more seriously.

3.3.1.4 Project-based approach

Unlike systems such as TQM and Taguchi methods, Six Sigma is usually carried out on a project-by-project basis. The spirit is still the same – continuous improvement – but the manifestation is different. With a project-based approach, a Six Sigma program can easily be identified and managed. A clear target must be specified in advance and examined to see whether a project should be carried out. Approved projects usually last between 4 and 6 months, and their performance is usually measured in terms of monetary returns.

3.3.1.5 Well-structured project team

Associated with the project-based approach is a well-designed project team structure. A Six Sigma project team consists of Champions, Master Black Belts, Black Belts and

Green Belts . The core of the operational Six Sigma team is made up of the Master Black Belts, Black Belts and Green Belts. Master Black Belts oversee Six Sigma projects and act as internal Six Sigma consultants for new initiatives. Black Belts are the core and full-time practitioners of Six Sigma.[5] Their main purpose is to lead quality projects and work full-time until projects are completed. They are also responsible for coaching Green Belts, who are employees trained in Six Sigma but spend only a portion of their time completing projects while maintaining their regular work role and responsibilities. This clear and comprehensive team structure makes the program tangible and manageable.

3.3.1.6 Clear problem-solving framework

Six Sigma provides a clear, systematic problem-solving framework, DMAIC, as the core of its technological base. Statistical tools, such as design of experiments, statistical process control and Monte Carlo simulations, and structured decision support tools, such as quality function deployment and failure mode and effects analysis, are integrated under this framework. The DMAIC approach has also been adopted for the service sector.[6] This approach mainly focuses on combating variation, the main root cause of quality problems. In addition, the design for Six Sigma (DFSS) framework offers a systematic means to address quality problems from the design phase of any product. All these provide clear, unambiguous, continuous frameworks for practitioners.

3.3.1.7 Systematic human resource development

Six Sigma emphasizes human resource development and calls for heavy investment in staff training. Practitioners of Six Sigma hold different titles such as Green Belt, Black Belt, Master Black Belt and Champion, which are related to the level of personal competency and roles in carrying out projects. Practitioners usually start from the more basic and applied Green Belt training. Then they proceed on to the next level of Black Belt to deal with problems in depth with more tools. Subsequently, their technical competencies will be elevated to those of Master Black Belt when they have gained the necessary technical and management experience for them to progress and effectively act as internal consultants to Six Sigma programs.

3.3.1.8 Project tied to bottom line

As pointed out, Six Sigma is implemented on the basis of projects. Once the key business processes are identified, every project will have a deadline and be tied to the bottom line. There is usually an audit of the newly improved way of operating processes, thereby enabling the company to assess the actual effectiveness of each project.

3.3.2 Weaknesses

3.3.2.1 High investment

A large amount of investment is required to train employees to be Green Belts, Black Belts, Master Black Belts and so on. It is generally recommended that an average of one

Black Belt be available per 100 employees.[1] Thus an organization of 10 000 employees would probably need to train 100 Black Belts.

Returns from Six Sigma may not be realized in the short term. There may be negative returns at first. Hence, companies wishing to embark on Six Sigma projects will have to master the philosophy, adopt the perspectives and maintain their commitment for an extended period. Because of this it may be difficult to justify the initial costs to stakeholders who have yet to see concrete results.

3.3.2.2 Highly dependent on corporate culture

The success of any Six Sigma implementation is very much dependent on the flexibility of the organization in being able to adapt its established functions and processes to the structured and disciplined Six Sigma approach. Six Sigma is not just a technically sound program with a strong emphasis on statistical tools and techniques, but also requires the establishment of a strong management framework.

In comparison with TQM models, Six Sigma places more emphasis on successful management elements. Thus a shift in the corporate culture within the organization is usually a necessity. This entails a shift in the internalized values and beliefs of the organization, ultimately leading to changes in internal behaviors and practices. This implies that if the company has an established and strongly traditional approach in its operations, changes are more difficult to carry out.

3.3.2.3 No uniformly accepted standards

There is as yet no general body for the certification of Six Sigma personnel or companies, though there are many diverse organizations issuing Six Sigma certificates. No unified standards and procedures have been set up and accepted so far. For companies considering building up a core Six Sigma expertise, the lack of a standardized body of knowledge and a governing body to administer them may result in varying levels of competency amongst so-called 'certified' Six Sigma practitioners. Every training organization determines its own training course content. Many of these training courses may be unbalanced in their focus or lack some critical methodological elements.

3.3.3 Opportunities

3.3.3.1 Highly competitive market and demanding customers

The current globalization and free trade agreements make the competition for market share more intense. Manufacturers are not competing locally or regionally, but globally. To gain or maintain one's market share requires much more effort and endeavor than ever before. Higher quality and reliability are no longer a conscious choice of the organization but a requirement of the market. For any organization to be successful, quality and reliability in the products that they offer have become one of the essential competing elements. This offers a great opportunity for Six Sigma since the

essence of Six Sigma is to achieve higher quality and reliability continuously and systematically.

3.3.3.2 *Fast development of information and data mining technologies*

Six Sigma depends heavily on data. Data measurement, collection, analysis, summarization and interpretation constitute the foundation of Six Sigma technology. Accordingly, data manipulation and analysis techniques play an important role in Six Sigma. Advanced information technology and data mining techniques greatly enhance the applicability of Six Sigma because modern technologies make data analysis no longer a complicated, tedious task. The availability of user-friendly software packages is certainly a good opportunity for the application of Six Sigma as it removes the psychological as well as operational hurdles of statistical analysis required in Six Sigma.

3.3.3.3 *Growing research interest in quality and reliability engineering*

The growing interest in quality and reliability engineering research represents another opportunity for Six Sigma. For example, research in robust design combined with Six Sigma produces an important improvement to Six Sigma – DFSS. While the traditional DMAIC approach mainly deals with existing processes, the more recent DFSS addresses issues mainly at the design stage; it introduces the idea of designing a process with Six Sigma capability, instead of transforming an existing process to Six Sigma capability. Interest in quality and reliability engineering research is growing and there is considerable potential for the improvement of Six Sigma.[7,8]

3.3.3.4 *Previous implementation of quality programs has laid foundation for the easy adoption of Six Sigma*

Modern quality awareness started in the mid-twentieth century. Since then, various quality programs have been developed and put in practice. These programs have laid a much needed foundation for the adoption of Six Sigma. Among these is TQM, which shares some similarities with Six Sigma, such as a focus on customer satisfaction and continuous improvement. Companies which have already implemented other quality programs generally find it less difficult to adopt Six Sigma, their previous experience serving as a valuable 'warm-up' exercise.

3.3.4 Threats

3.3.4.1 *Resistance to change*

The success of Six Sigma requires cultural change within the organization.[9,10] Six Sigma should be embraced in the organization as a corporate philosophy rather than as 'yet another' quality initiative. Six Sigma revolutionized the way organizations work by introducing a new set of paradigms. Although Six Sigma tools are not difficult to learn, the managers and the rest of the workforce who have been with the organization for a long time might view them as an additional burden. These managers rely mainly

on their experience in dealing with problems and are confident enough to use their intuition rather than resort to statistical tools deriving information from available data. Such an attitude may be harmful to the success of Six Sigma. The middle managers and supervisors who have experienced many other quality initiatives may regard Six Sigma as yet another transient offering that will pass in due course.

3.3.4.2 Highly competitive job market

Few companies practice life-long employment in today's competitive job market. This is even more prevalent given the rapidly changing economic, social and technological environment. People tend to change jobs more frequently, whether it be in pursuit of 'better prospects' or involuntarily. The impact of frequent job changing is further worsened by the fact that appreciable benefits from serious Six Sigma work can only be felt a few years after projects have been initiated.

Corporate leadership plays a vital role in the successful implementation of Six Sigma. The implementation structure of Six Sigma demands strong support from the Champions or executive management.[11] Any changes in the executive management will have adverse effects on the implementation. With hostile market conditions, corporate leadership has become relatively more volatile. Chief executives are changed frequently, or changes may be brought about through mergers and acquisitions between organizations. When higher-level management is changed frequently, it may be difficult to maintain the same level of top-down commitment to Six Sigma initiatives.

3.3.4.3 Cyclical economic conditions

Economic trends are usually cyclical. In good times, companies may be more willing to invest additional income on process improvement efforts. This tendency may be reversed during situations of economic downturn as companies struggle to keep afloat. Such situations may be unhealthy for Six Sigma implementation, again particularly in view of the fact that an extended training and application phase is often needed before significant financial gains can be seen.

3.3.5 Summary of assessment

A consolidated table of the SWOT analysis is shown in Table 3.1.

3.4 FURTHER THOUGHTS

Six Sigma has been at the forefront of the quality movement in recent years. However, the hyperbole that often accompanies claims as to its effectiveness could lead to unrealistic expectations. For example, it is commonly claimed by Six Sigma consultants that a typical Six Sigma project could save a company at least US$200 000. A specific Six Sigma project might well achieve this, but what happens when Six Sigma is adopted by an organization that has an annual turnover of, say, US$100 000? Blindly marketing Six Sigma with unrealistic promises will only hurt Six Sigma in the long

Table 3.1 Consolidated matrix of SWOT analysis on Six Sigma strategy.

	Positive factors	Negative factors
Internal	1. Customer focus 2. Data-driven and statistical approach to problem solving 3. Top-down support and corporate-wide involvement 4. Well-structured project personnel and teams 5. Clear problem-solving framework (DMAIC) 6. Project-based, result-oriented approach 7. Systematic human resource development 8. Performance tied to bottom line	1. Heavy investment 2. Highly dependent on corporate culture (receptiveness to change) 3. No uniformly accepted standards
External	1. Highly competitive market and demanding customers 2. Fast development of information and data mining technologies 3. Growing research interest in quality and reliability engineering 4. Previous implementation of quality programs has laid foundation for adoption of Six Sigma	1. Resistance to change 2. Highly competitive job market 3. Cyclical economic conditions

run, as business leaders are usually hard-nosed individuals with more fascinating stories to tell than a shallow Six Sigma vendor. Thus for organizations that have not yet adopted Six Sigma for quality and business excellence, Six Sigma tends to be more credible if its strengths and weaknesses, together with potential opportunities and threats, can be laid out objectively for the appreciation of their senior managers. The SWOT format is a concise way of highlighting the essential points.

It may also be remarked that among those already in the know, the understanding of Six Sigma varies from organization to organization. Some regard it as a management philosophy and some take it as a well-designed statistical package. Certainly, the interpretation that would lead to beneficial exploitation of Six Sigma's full potential is to view it as both. Two of the key elements of Six Sigma successes are commitment from top management and the corporate culture. If the top management is highly committed and the corporate culture is dynamic and receptive to change, Six Sigma can be used as a strategic guideline for better financial returns and business excellence. At the operational level, Six Sigma projects will not create an impact if the requisite

statistical concepts and tools are not effectively applied – therefore mastery of technical details is essential.

Finally, it must be pointed out that 'there are Black Belts and Black Belts', just as 'there are consultants and consultants'. It would be a gross mistake to assume that all persons carrying a Black Belt certificate have the same capabilities. One should bear in mind that a poor project is not necessarily due to the ineffectiveness of Six Sigma, but could well be the result of the inadequate capability of the Black Belt in question. Indeed, this would be a good subject for serious researchers of Six Sigma, since up until now the role played by Six Sigma professionals – Black Belts, Green Belts, etc. – has only been *prescribed* in various forums; in the light of twenty years of Six Sigma in the corporate world, the time is ripe for a formal study of its real effectiveness by either technical or business researchers.

REFERENCES

1. Harry, M. and Schroeder, R. (2000). *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations*. New York: Doubleday.
2. Pande, P.S., Newman, R.P. and Cavanagh, R.R. (2000) "*The Six Sigma Way*. New York: McGraw-Hill.
3. Pyzdek, T. (2001) Why Six Sigma is not TQM. *Quality Digest*, February. http://www. qualitydigest.com/feb01/html/sixsigma.html.
4. Howell, D. (2001) At sixes and sevens. *Professional Engineering*, 14(8), 27–28.
5. Hoerl, R.W. (2001) Six Sigma Black Belts: What do they need to know? *Journal of Quality Technology*, 33(4), 391–406.
6. Goh, T.N. (2002) A strategic assessment of Six Sigma. *Quality and Reliability Engineering International*, 18, 403–410. See also Chapter 2, this volume.
7. Montgomery, D. (2001) Beyond Six Sigma. *Quality and Reliability Engineering International*, 17(4), iii–iv.
8. Goh, T.N. (2005) *Quality Engineering Applications of Statistical Design of Experiments*. Singapore: John Wiley (Asia).
9. Hendricks, C.A. and Kelbaugh, R.L. (1998) Implementing Six Sigma at GE. *Journal for Quality and Participation*, 21(4), 48–53.
10. Blakeslee, J.A. Jr (1999) achieving quantum leaps in quality and competitiveness: Implementing the Six Sigma solution in your company. *Annual Quality Congress Transactions*, 53, 486–496.
11. Henderson, K.M. and Evans, J.R.(2000) Successful implementation of Six Sigma: Benchmarking General Electric Company. *Benchmarking: An International Journal*, 7(4), 260–282.

4

The Essence of Design for Six Sigma

L. C. Tang

Design for Six Sigma has been proposed as a new breakthrough strategy for organizations involved in product design and development. It has also been adopted for the design and improvement of transactional processes. In this chapter, we present an outline of the IDOV roadmap and the associated deliverables at each phase. Brief descriptions of selected tools and concepts, such as QFD and reliability, which are not covered in the traditional DMAIC training curriculum, are included. A table summarizing the quality tools and their respective purposes is also presented.

4.1 INTRODUCTION

Design for Six Sigma (DFSS) was coined by General Electric (GE) in 1996 when breakthrough improvements had to be derived from the design of their CT scanners rather than from the production line in order to fulfill the much higher reliability requirements of customers.[1] This is indeed a natural progression for companies engaging in Six Sigma. After plucking most of the low hanging fruit, improvements must be sought in upstream processes, particularly the design process. Excellence in manufacturing has its limitations as it can never eradicate inherent design weaknesses. Product design imposes constraints on production processes. DFSS is thus part of Six Sigma deployment strategies that start with manufacturing and permeate through design, marketing, and administrative functions in a vertically integrated enterprise.

The DMAIC quality improvement framework was found to be inadequate as it was originally conceived for operational activities. Two common frameworks have since been adopted: IDOV (identify, design, optimize, and validate) and DMADV (define,

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

measure, analyze, design, and verify). In this chapter we present a general outline of the IDOV framework proposed by GE and implemented in Seagate Technology since the turn of the millennium. Since its adoption in late 1990s, successes have been reported by these corporations. Notably, GE Medical has secured leadership in terms of market share worth some US$2 billion globally for its CT scanners; and Seagate Technologies was presented a global Six Sigma Excellence Award in 2004 for "Best DFSS Project of the Year". It is thus beneficial for other Six Sigma believers to understand the DFSS design/development process and learn the major tools in DFSS deployment.

4.2 THE IDOV ROADMAP

The primary objective of DFSS is to design a product that pleases customers and achieves the maximum possible profitability within the safety constraints. To do so, it is imperative to manage risk during the product development process by assessing and controlling it by exercising trade-offs. This can only be achieved through a structured framework that emphasizes systems thinking and a toolset that allows the impact of decisions taken upstream on downstream operations to be quantified. These prerequisites motivate the IDOV roadmap for DFSS. The key phases are described in the rest of this section.

4.2.1 Identify

This is a highly strategic phase as one needs to identify the intended market segment for the product under design. On a macro scale, a cross-functional team must have a good idea of the emerging technologies, the market trend and the areas of growth. At the same time, the team must have their feet on the ground to understand their customers' needs. With the aid of quality function deployment (QFD), the team will then translate critical-to-quality (CTQ) requirements into objective measures which, in turn, give the design targets and specifications. Using the target values, a gap analysis looking into various technological and functional aspects such as product performance, the stakeholders and the competitive advantage, complemented by a design failure mode and effects analysis (FMEA), can be carried out. The final deliverable is a strategic plan that consists of three major elements: a project charter, the level 1 House of Quality, and matrices from gap analysis.

The *project charter* is a first document that sets the stage for the project. It should be clear and concise in order to get the project off the ground, and will be revised along the way. It will be used to facilitate communication and obtain approval and buying-in from the sponsor and stakeholders. The elements of a project charter are listed in Table 4.1.

QFD basically involves several steps which aim to translate the voice of the customer into relevant targets and specifications for the technical attributes of a product, its key components, the manufacturing processes, and the related operations through a series of flow-down matrices commonly known as the House of Quality.[2] An example of a level 1 House of Quality is presented in Table 4.2.

Table 4.1 Elements of a project charter.

Elements	Content
Project title	A concise phrase for the intended work to create an identity.
Team members	List the team members and identify the leader. This is to establish the roles and responsibilities of team members.
Purpose	Give a statement of purpose in a concise manner so that a clear common vision is established and communicated. It must relate to the bottom line so as to establish a business case.
Scope	Outline the expected tasks to be included so that the milestones are clearly defined to establish interim targets and to avoid unnecessary project creep.
Goal	List the final deliverables and define the metrics to measure success. This gives a specific target to aim for.
Plan and time line	Identify the resources – manpower, equipment, etc. – needed over the duration of the project and the deadline for each milestone defined in the scope list. A simple Gantt chart may be included.

One of the key inputs into QFD is the Kano classification of customers' needs listed in the row entries. Three major classifications are as follows:

- *Attractive elements* (ATT) are 'delighters'. When delighters are absent, the customer is neutral (not dissatisfied).
- *Must-be elements* (MB) are element which, if absent, will cause customers to be very dissatisfied.

Table 4.2 An example of a level 1 House of Quality for a printer.

				1	2	3	4	5	6	7	8	9	10
HOW / WHAT		Kano Class	INDEX	Resolution (Dpi)	Print speed (page/min)	power (W)	Cooling fan speed (rpm)	noise level (dB)	duplex feature	Long toner life (page)	Cost ($)	reliability (MTTF)	Maintainability (MTTR)
1	Clarity	MB	5	■	▤					▤	▤		
2	fast	OD	4	▤	■			▤	□		▤		
3	lasting	OD	4								□	■	
4	low maintenance cost	OD	4	▤						■			▤
5	low power consumption	OD	3			■							
6	low noise	OD	3				▤	■	□		□		
7	paper saving	ATT	2						■		▤		
8	Reasonable price	MB	5	▤	▤					▤	■	▤	▤
	Raw Weight			84	66	27	9	39	25	66	85	51	27
	Weight %			18	14	6	2	8	5	14	18	11	6
	Weight 1-5 Scale			5	4	2	1	2	1	4	5	3	2

RELATIONSHIPS		
■	STRONG	9
▤	MEDIUM	3
□	WEAK	1

Total	Max
479	85

- *One-dimensional elements* (OD) are those where there is a strong correlation between satisfaction levels and the degree of presence or absence of a feature or service.

The general rule of thumb is to fulfill all MB needs and to establish leadership position in certain OD needs in the market. The ATT needs are good-to-have features and can be used to differentiate a product from others. It is worth noting that some ATTs may later evolve into future MB needs.

In essence, the QFD matrix can be completed by taking the following steps:

1. List the customers' needs (i.e. the whats).
2. Classify and assign priority index to the customers' needs using the Kano model.
3. List the technical attributes of the design to fulfill the needs (i.e. the hows).
4. Determine the relationships between the whats and the hows; strong relationships are given a score of 9, moderate a score of 3, and weak a score of 1 (see Table 4.2).
5. Calculate scores for each how. This is done by multiplying the priority index column with the relationship score under the respective how column. For example, under 'resolution' in Table 4.1, the raw weight is given by $5 \times 9 + 4 \times 3 + 4 \times 3 + 5 \times 3 = 84$. The raw weights are then normalized to a scale of 1 to 5 by dividing each raw weight by the maximum value.
6. Define the targets and limits on the hows (not shown).
7. Perform technology, measurement, and competitive analyses (not shown).
8. Determine the course of action.
9. Flow down requirements.

QFD is a very useful tool to facilitate and to document the thinking process in fulfilling customer needs. It also provides a common ground for teamwork and triggers subsequent actions in satisfying the needs. During the product development process, information not only flows down to the lower-level matrix but also flows up in order for certain design limits to be revised to achieve the desired trade-off in manufacturing cost constraints and customer satisfaction. There has been some work on performing such trade-off analysis; notably, Tang and Paoli proposed a spreadsheet model using linear programming to obtain the Pareto optimal solutions.[3]

4.2.2 Design

While the identify phase is mainly strategic in nature, the design phase requires major engineering effort to be integrated with statistical skills, as this is where conceptual design takes place and a feasible design that focused on customers' requirements emerges. Compared to the traditional DMAIC roadmap, the tools in the design phase are a subset of the most powerful toolset from the measure, analyze and improve phases of DMAIC where the transfer function between the key process input variables (KPIVs) and key process output variables (KPOVs) is developed for product characterization rather than for process optimization purposes. Some advanced concepts and techniques will be presented in the later chapters.

The key tasks in the design phase include the following:

1. Formulate conceptual designs and develop into possible design solutions.
2. Evaluate alternative designs and quantify their associated risk.

3. Understand and verify the capability of the measurement system.
4. Identify KPIVs and KPOVs for each CTQ.
5. Establish the relationship between KPIVs and KPOVs.

The main deliverable is the transfer function between the KPIVs and KPOVs which must be of adequate predictive power so that the associated design risk can be quantified with high level of confidence. We will present some important concepts and tools in Part D of this book. An excellent reference on this topic is the book by Wu and Hamada.[4]

4.2.3 Optimize

In the optimize phase, armed with the transfer functions for major CTQs, appropriate tolerances will be established using Monte Carlo simulation. To carry out the intended simulation, assumptions of input distributions for the KPIVs are needed so that the KPOVs can be simulated using the transfer functions. For example, suppose that the KPOV (Y) of a CTQ can be represented by the function

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_1 X_2,$$

where the X_i are the KPIVs. To assess the response, the input distributions of the X_i are assumed and a large number of the response of the KPOV can then be simulated. This can easily be done in software such as MINITAB or even Excel. The capability measures with respect to the specification developed in the identify phase can then be assessed. Monte Carlo simulation also provides a convenient means to perform sensitivity analysis by changing the parameters of the input distributions and the distributional assumptions. The process can be repeated to optimize the inputs and obtain realistic tolerance.

Usually the number of replications – the number of random variables generated each time for a particular KPIV – is very large, say, 10 000. Correspondingly, the number of responses is also very large. It is thus not advisable to try fitting a distribution to the data as "all models are wrong". Conclusions should be drawn directly from the empirical distribution or the histogram of the response. The resolution of the results can easily be improved by increasing the number of replications.

Another major task during this phase is to estimate the reliability of the product by carrying out some accelerated tests which involve testing the product under higher stress levels than those under which it will usually operate. The key to success is to identify the correct stressor or way to shorten the lifetime of the product so that failure data for the same failure mode as those under usual operating conditions can be obtained within reasonable test duration. The best place to start looking for the correct way of stressing the product is the design FMEA table drafted in the identify phase as it lists major failure modes and how failures are triggered. Field data of previous generations of the product must also be examined in order to understand how the product fails, to identify critical failure modes and to set realistic reliability targets. A detailed treatment of this topic is beyond the scope of this chapter, and readers may refer to Tobias and Trindade for an excellent treatment of the subject.[5] Here, we present the essential ideas by assuming that the failure time follows an exponential distribution with mean time to failure θ (or failure rate $\lambda = 1/\theta$).

A typical desired outcome of a reliability test is to claim that, with $100(1-\alpha)\,\%$ confidence, at most $p\,\%$ will fail by a certain time τ. Statistically, the largest value for τ is the lower $100(1-\alpha)\,\%$ confidence limit of the pth percentile of the time to failure. For exponential failure time, this is given by

$$TTF_{p,\mathrm{LCL}} = \frac{2r\hat{\theta}}{\chi^2_{\alpha,2r+2}}\left(-\ln(1-p)\right),$$

where r is the number of failures in a time-censored test, $\hat{\theta}$ is the estimate of θ and $\chi^2_{\alpha,2r+2}$ is the upper αth percentile of a chi-square statistic with $2r+2$ degrees of freedom. If a reliability specification in terms of τ is given, then the lower confidence limit obtained from the test must be more than τ. Since $\hat{\theta} = TTT/r$, where TTT is the total time on test, we have

$$TTT \geq -\frac{\tau\,\chi^2_{\alpha,2r+2}}{2\ln(1-p)}.$$

As a result, a high reliability requirement in terms of large τ or small p will require a long test time and large sample size. For the same product tested under higher stress, the same reliability requirement will translate into smaller τ and the test duration can be shortened accordingly. In other words, a time compression can be achieved, commonly refered to as the acceleration factor. This factor can be computed from the model that relate the time to failure to the stress level. For example, the most common model when the applied stress is temperature is the Arrhenius model, and the acceleration factor is given by

$$AF = \exp\left[\frac{E_\mathrm{a}}{k}\left(\frac{1}{T_0}-\frac{1}{T_\mathrm{s}}\right)\right].$$

where E_a is the activation energy, k is the Boltzmann constant ($8.617\times10^{-5}\,\mathrm{eV\,K^{-1}}$), and T_0 and T_s are the temperature in Kelvin under use and stress condition respectively.

Putting these together, the total test time at higher stress level is then given by

$$TTT \geq -\frac{\tau\,\chi^2_{\alpha,2r+2}}{2AF\,\ln(1-p)}.$$

A simple reliability test plan that specifies the sample size and the test condition (say, higher temperature) can then be derived from the above expression.

Subsequent to setting the tolerance, checking feasibility, and performing capability and reliability assessment and the associated sensitivity studies, the result will be verified against customers' requirements. The final deliverables are a set of well-conceived tolerance setting, the reliability and capability scores.

4.2.4 Validate

The main tasks of this phase are to conduct confirmatory runs with appropriate sample sizes and to prepare the design for production. The purpose of validation of the performance of the selected design is to ensure that the key performance indices are achievable under mass production. After validation, control plans will be put in place to ensure that the KPOVs of the design are within acceptable limits. For example, to validate a target process capability index (C_{pk}), an approximate sample size formula

for showing that C_{pk} is greater than say, $C_{pk0} - \Delta$, where C_{pk0} is the true value, is given by

$$n = \frac{(Z_\alpha + Z_\beta)^2 (9C_{pk0}^2 + 2)}{18\Delta^2}.$$

The above approximation works well for large $n(>50)$ and $C_{pk0} > 1$, as it is based on an asymptotic result for C_{pk}.

A consolidated table summarizing the DFSS tools and their respective purposes is given in Table 4.3. For a seasoned Black Belt who is conversant with DMAIC tools, acquiring three additional tools, namely, QFD, design for reliability, and Monte Carlo

Table 4.3 Six Sigma tools and their objectives in DFSS.

	Quality tools	Purpose
Identify	• Project management tools such as Gantt charts and critical path analysis • Voice of customers • QFD; FMEA, Benchmarking; Stakeholder analysis; leverage chart; Pareto charts	• To develop a project charter • To identify customers • To translate customers' wants into product design requirements; requirements into specifications and to perform gap analysis
Design	• Brainstorming; affinity diagram; reality trees • TRIZ (theory of inventive problem solving); cause-and-effect analysis • Gage R&R • Statistical thinking; *t*-, *F*-test, ANOVA • Screening experiments • Robust design • Response surface method, multiple regression	• To formulate design concepts • To brainstorm conceptual design • To get measurement systems ready for design evaluation • To evaluate design options • To identify KPOVs and KPIVs • To evaluate design sensitivity to environment changes • To develop transfer functions
Optimize	• Monte Carlo simulation • Capability analysis • Design for reliability • Two-level fractional factorials • DFSS score card	• To perform statistical tolerancing • To analyze and optimize capability measures • To develop reliability metrics and to conduct accelerated testing • To demonstrate feasibility and conduct sensitivity analysis • To review current design achievement
Validate	• Sample size determination • Statistical process control • Poka-yoke(mistake proofing)	• To conduct confirmatory runs • To design control plan and develop control charting tools • To error proof design and processes

simulation for statistical tolerancing, would enable him to be operationally ready for DFSS.

4.3 THE FUTURE

The DFSS roadmap is still evolving. For example, in April 2004, the Six Sigma corporate vice-president of GE (the originator of DFSS), Piet van Abeelen, announced that "The roadmap for DFSS at GE is now DMADOV, with measurements and optimization steps strictly enforced. Because of the nature of the newly re-defined DMADOV, it is used for new product introduction or large enterprise-wide processes." New systems engineering methodologies for DFSS are still being proposed and tested in practice. Other multinational corporations such as Philips and Siemens are still trying to adapt the principles of DFSS to their existing product development process. It is thus natural for the roadmap to keep evolving over time as innovation is the key driving force in any design-oriented enterprise.

REFERENCES

1. Harry, J.M. and Schroeder, R. (2000) *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations.* New York: Doubleday.
2. kao, Y. (1990) *Quality Function Deployment: Integrating Customer Requirements into Product Design.* Cambridge, MA: Productivity Press.
3. Tang, L.C. and Paoli, P. (2004) A spreadsheet-based multiple criteria optimization framework for Quality Function Deployment. *International Journal of Quality and Reliability Management*, 21(3), 329–347.
4. Wu, C.F.J. and Hamada, M. (2000) *Experiments: Planning, Analysis, and Parameter Design Optimization.* New York: John Wiley & Sons, Inc.
5. Tobias, P.A. and Trindade, D.C. (1995) *Applied Reliability*, 2nd edition. New York: Van Nostrand Reinhold.

5

Fortifying Six Sigma with OR/MS Tools

L. C. Tang, T. N. Goh and S. W. Lam

Six Sigma as a quality improvement framework cannot remain static if it is to sustain its value for businesses beyond the first wave of applications. This chapter explores the possibility of enhancing the usefulness and effectiveness of Six Sigma by the integration of established operations research/management science (OR/MS) techniques, focusing on operational and transactional environments. A matrix relating the deliverables to an integrated Six Sigma toolset fortified with new OR/MS tools under the DMAIC roadmap for a hybrid of operational and transactional environments is then presented. A case study is discussed to demonstrate the usefulness of one of the proposed OR/MS tools.

5.1 INTRODUCTION

Six Sigma is a systematic, highly disciplined, customer-centered and profit-driven organization-wide strategic business improvement initiative that is based on a rigorous process-focused and measurement-driven methodology. Six Sigma makes use of sound statistical methods and quality management principles to improve processes and products via the define–measure–analyze–improve–control (DMAIC) quality improvement framework to meet customer needs on a project-by-project basis. With many high-profile adoptions by companies such as General Electric (GE) in the 1990s, Six Sigma had spread like wildfire by the end of the twentieth century.[1]

This chapter is based on the article by L. C. Tang, T. N. Goh, S. W. Lam and C. W. Zhang, "Fortifying Six Sigma with OR/MS Tools", *Quality and Reliability Engineering International*, to be published, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd.

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

Operations research/management science (OR/MS) techniques have been extensively applied to a wide range of areas, including transportation, manufacturing, telecommunications, the military, financial planning, public services, and health care. They are effective tools for improving the efficiency and productivity of organizations. OR/MS techniques, as implied by the name, are concerned with the conduct and improvement of 'operations' as well as the practical management of organizations. Another important characteristic of OR/MS is that, rather than simply improving the status quo, its primary goal is to identify a best possible course of action.[2] This is also a fundamental goal of all Six Sigma projects, commonly exemplified by the Improve phase. Consequently, it is quite natural to integrate OR/MS techniques into the framework of Six Sigma.

As an approach to organizational excellence, Six Sigma as it has been practiced can certainly be enhanced.[1] In spite of its success, Six Sigma can be made better with OR/MS techniques. For example, mathematical programming and decision analysis techniques can improve project selection and planning; and queuing and simulation techniques can improve the analysis and operations of transactional systems. Mathematical programming techniques are also instrumental and essential in solving various optimization problems throughout the implementation of Six Sigma.

This chapter attempts to explore possibilities of further enhancing the usefulness and effectiveness of Six Sigma via the integration of OR/MS techniques into Six Sigma deployment. A new roadmap for Six Sigma Black Belt (BB) training which contains an expanded curriculum that includes new OR/MS techniques is proposed. A case study to improve the efficiency of processes in the transactional environment through more effective allocation of human resources is then presented to illustrate the usefulness of OR/MS techniques within the proposed roadmap.

5.2 INTEGRATION OF OR/MS INTO SIX SIGMA DEPLOYMENT

Six Sigma is not only a performance target that applies to a single critical-to-quality (CTQ) characteristic,[3] but has evolved into a strategic business improvement approach that seeks to find and eliminate causes of mistakes or defects in business processes by focusing on outputs that are of critical importance to customers.[4] The ultimate goal of Six Sigma is business improvement, particularly in terms of bottom-line results, customer satisfaction and market share.

Most Six Sigma tools are statistical methods and quality management tools, such as design of experiments (DOE), response surface method (RSM), robust design, statistical process control (SPC), quality function deployment (QFD), failure mode and effect analysis (FMEA), capacity analysis, hypothesis testing, analysis of variance (ANOVA), and regression analysis. None of these tools are not new; it is their integration that is unique and an important factor in the success of Six Sigma. Six Sigma is characterized by integrating a collection of statistical methods and quality improvement tools in a purposeful, systematic and sequential manner to achieve synergistic results far exceeding what is possible with the isolated application of single tools or methods. This is an important feature of Six Sigma. It is not a simple 'summation' of a bunch of tools but the systematic integration of these tools for problem-solving that produces breakthrough results.

Everything has a life cycle. Although it can work very well, Six Sigma cannot be a static framework, if the success is to last.[5] Despite its current success, Six Sigma has its inherent limitations and cannot be a universal solution for every organization in every situation.[1] The standard Six Sigma tools and methods are effective in dealing with quality-related issues. However, statistical methods and quality management tools are not sufficient for tackling all business improvement related problems. For example, the existing Six Sigma tools and methods are inadequate for dealing with problems such as production and service planning and scheduling, inventory control and management, supply chain management, operations scheduling, workforce scheduling, and so on. Moreover, quality management tools such as QFD can be greatly enhanced by OR/MS tools,[6] and it is becoming apparent that the standard training package received by BBs is not sufficient for dealing with complex process-improvement projects. In fact, all these problems are closely related to the overall performance of an organization.

Customer satisfaction is a reflection on the state of the business and depends on three things: delivering a defect-free product or service; delivering a product or service on schedule; and delivering a product or service at the lowest possible cost.[3] Though standard Six Sigma tools and methods are effective in handling the first problem, they are not able to solve the last two. OR/MS techniques are well positioned to effectively solve these problems, among others. Many techniques used by OR/MS practitioners could and should be integrated into Six Sigma applications to complement the existing standard Six Sigma tools. In fact, common concepts and procedures of operations management and project management can play an extremely useful role in Six Sigma projects. Many Six Sigma success stories have involved the use of OR/MS techniques. It could be loosely argued that every OR/MS technique can be used in the deployment and implementation of Six Sigma.

One of the most useful purposes of applying OR/MS techniques is for 'improvement'. Consequently, various OR/MS techniques are well fitted into the improve phase in Six Sigma deployment. The Define phase also involves scores of problems, such as project selection and planning, production and service planning, training and education planning, resource allocation, investment decision-making, and facility and service layout and location, which conventional Six Sigma tools cannot handle but OR/MS techniques can. In the Control phase OR/MS techniques can also be applied, for example, to optimize the design of control charts and control schemes, and to improve maintenance management.

Considering the fact that Six Sigma programs adhere strictly to a systems perspective on quality improvement, it is quite natural to observe the trend of integrating Six Sigma with other business-improvement tools and methods such as lean manufacturing, design for Six Sigma (DFSS), and supply-chain operations reference (SCOR).[7] In the long run, to really accomplish the simultaneous objectives of Six Sigma and lean enterprise, practitioners will have to gain not only a solid understanding of additional statistical tools, but also knowledge of industrial engineering and OR techniques, such as systems simulation and factory modeling, mathematical optimization methods, and queuing networks.[6] Although these techniques have been included in Master Black Belt(MBB) traning, the number of MBBs is far too small in a typical enterprise to have an impact. Future successes of Six Sigma can only be brought about by dedicated teams of BBs mastering a set of synergistic tools arranged in a compact and logical sequence for problem-solving. The next section presents a new roadmap for a BB training program.

5.3 A NEW ROADMAP FOR SIX SIGMA BLACK BELT TRAINING

In developing the new training program, we first compare and contrast the training needs in different environments. In Table 5.1, we present an expanded curriculum based on a typical Six Sigma BB training program in an operational environment alongside a new curriculum for Six Sigma BBs in a transactional environment. Note that only the basic OR/MS techniques have been incorporated in the new curriculum so that these topics could be covered within the usual 4-week BB training program. Table 5.2 summarizes all the supplementary OR/MS techniques extracted from Table 5.1.

Six Sigma is a process-focused quality-improvement initiative. The 'processes' in manufacturing/operational and transactional environments are somewhat distinct and thus demand partially different toolsets during the implementation of Six Sigma as well as in a BB training program. From Table 5.1, it can be seen that the major difference between the manufacturing/operational and transactional roadmap is in the Analyze and Improve phases, with slight differences in other phases. OR/MS techniques such as forecasting, queuing, simulation and modeling are essential tools in the Analyze phase since system-level analysis is usually needed in a transactional environment. In the Improve phase, major tools used in manufacturing/operational environment are DOE techniques; in contrast, queuing and mathematical programming techniques are usually needed for transactional environments. From Table 5.1, OR/MS techniques appear to be much more applicable in a transactional environment. However, it should be noted that Six Sigma BBs working in manufacturing sectors are also expected to tackle transactional issues. This underscores the importance and necessity of integrating OR/MS techniques with Six Sigma. The current evolution of Six Sigma is not simply a transition from the original manufacturing sectors to service sectors but a vehicle for making deep cultural change, inculcating system thinking and problem-solving, leading to quantifiable benefits.

5.3.1 Basic OR/MS techniques

5.3.1.1 Decision analysis

Various decision-analysis techniques are useful tools for making 'good' decisions involved in Six Sigma deployment as well as other business operations. Effectively made decisions have a profound impact on overall business performance. Multiobjective decision-analysis techniques can be used in, for example, Six Sigma projects, material, vendor, product and process selection. Multiobjective decision-analysis techniques are also useful tools to assist organizations' strategic and tactical decision-making. Meanwhile, sensitivity analysis is usually done in conjunction with decision-analysis to assess the sensitivity of the decisions made with uncertain factors. While decision-analysis techniques could be applied in the whole process of Six Sigma deployment as each phase may entail some decision-making, its major role is in the Define phase of the integrated Six Sigma roadmap discussed in Section 5.3.2.

5.3.1.2 Mathematical programming

Mathematical programming techniques include various mathematically rigorous optimization tools such as linear programming, integer programming, mixed integer

Table 5.1 An expanded list of Six Sigma tools.

	Manufacturing/operational environment	Transactional Environment
Define	Project selection	Project selection
	Probabilistic risk thinking and strategic planning	Probabilistic risk thinking and strategic planning
	Decision analysis	Decision analysis
	Process mapping	Process mapping
	Project management tools	Project management tools
Measure	QFD and Kano analysis	QFD and Kano analysis
		Gap analysis
	Sampling (data quantity and data quality)	Sampling (data quantity and data quality)
	Measurement system analysis	Measurement system analysis
	SPC Part I (concepts, implications of instability)	
		Run charts (or time series graphs)
	Capability analysis	Capability analysis
	Monte Carlo simulation and statistical distributions	
Analyze	Basic graphical improvement tools	Basic graphical improvement tools
	FMEA	FMEA
	Hypothesis testing	Hypothesis testing
	Confidence intervals	
	ANOVA	ANOVA
	Correlation and regression analysis	Correlation and regression analysis
	Reliability models and measures	
		Cost analysis
		Forecasting
		Basic queuing systems
		Simulation and modeling
Improve	DOE (factorial, fractional factorial, blocking, nested and RSM)	DOE (factorial, fractional factorial and blocking)
	Robust design	
		Optimization and control of queues
		Mathematical programming techniques
		Heuristics
	Sensitivity analysis	Sensitivity analysis
Control	Mistake proofing	Mistake proofing
	Validation testing	Validation testing
	Control plans	Control plans
	SPC Part II: Control charts	Basic control charts

programming, nonlinear programming, network programming, dynamic programming, goal programming, multiobjective mathematical programming, and stochastic programming. Problems selected for Six Sigma projects are not limited to the field of engineering but cover quality issues in transactional, commercial and financial areas as well, with an explicit and strong customer focus.[8] Mathematical programming

Table 5.2 A summary of OR/MS techniques integrated into Six Sigma phases.

	OR/MS tools
Define	Mathematical programming techniques for resource allocation and project selection
	Decision analysis
	Project management tools
Analysis	Forecasting
	Basic queuing systems
	Simulation and modeling
Improve	Optimization and control of queues
	Mathematical programming techniques
	Heuristics

techniques, sometimes in conjunction with sensitivity analysis, can be exploited to solve such problems. These techniques have been predominantly used in production planning and operations management. They can be deployed in Six Sigma projects for project selection and planning during the Define phase of the Six Sigma deployment for selecting an optimal number of projects or to achieve profit maximization or cost minimization goals in general. Problems such as Six Sigma resources allocation, Six Sigma facilities layout and location, and production and service planning can also be solved using mathematical programming techniques. These applications may take a wide variety of forms depending on the particular problem situation and the various objectives involved. For example, given some limited capital budget, the decision of how to select a subset of proposed Six Sigma projects to invest in can be readily modeled as a single or multiobjective knapsack problem. Solution techniques for problems of this type are discussed by Martello and Toth,[9] and by Zhang and Ong,[10] among others.

Besides the Define phase, applications of mathematical programming techniques are interspersed in all subsequent phases. In particular, as the objective of mathematical programming techniques is optimization, various techniques can naturally be weaved into the Improve phase to solve various optimization problems. For example, a general framework for dual response problem can be cast using multiobjective mathematical programming.[11,12] Nonlinear optimization techniques can be applied, for example, to optimize mechanical design tolerance[13] and product design capability,[14] as well as to estimate various statistical parameters. In the Control phase, nonlinear optimization techniques have been applied to optimize the design of control charts, including economic design, economic-statistical design and robust design, design of sampling schemes and control plans. Examples of these applications can be found in many papers[15–29]. Some of these techniques are included in the proposed Six Sigma roadmap discussed in Section 5.3.2.

In addition, heuristics, the most popular ones of which include the classical metaheuristics of simulated annealing, genetic algorithms and tabu search, are a class of effective solution techniques for solving various mathematical programming and combinatorial optimization problems, among others. It is thus proposed that a brief introduction to heuristics should also be included in the training of Six Sigma BBs and

the deployment of Six Sigma, particularly in the Improve phase. Detailed treatment, however, can be deferred to an MBB program.

5.3.1.3 Queuing

Queuing theory is concerned with understanding the queuing phenomenon and how to operate queuing systems in the most effective way. Providing too much service capacity to operate a system incurs excessive costs; however, insufficient service capacity can lead to annoyingly long waiting times, dissatisfied customers and loss of business. Within the context of business-improvement, queuing techniques have frequently been applied to solve problems pertaining to the effective planning and operation of service and production systems. Specific application areas include service quality, maintenance management, and scheduling. Queuing techniques have been widely applied in such areas as manufacturing, service industries (e.g. commercial, social, healthcare services), telecommunications, and transportation. Queuing techniques can play a useful role in Six Sigma deployment, particularly in analyzing and improving a system providing services.

5.3.1.4 Simulation and modeling

Simulation is an exceptionally versatile technique and can be used (with varying degrees of difficulty) to investigate virtually any kind of stochastic system.[2] For instance, simulation can help to improve the design and development of products as well as manufacturing and service processes for a wide variety of systems (e.g. queuing, inventory, manufacturing, and distribution). Simulation has been successfully deployed in DFSS to replace costly preliminary prototype testing and tolerancing. Also, simulation provides an attractive alternative to more formal statistical analysis in, for example, assessing how large a sample is required to achieve a specified level of precision in a market survey or in a product life test.[21] Bayle *et al*.[22] reported the approach of integrating simulation modeling, DOE and engineering and physical expertise to successfully design and improve a braking subsystem that would have not been accomplished by any individual tool or method alone.

For system operations analysis, simulation is an indispensable companion to queuing models as it is much less restrictive in terms of modeling assumptions.[23] Queuing and simulation techniques also play important roles in inventory control[24] and supply chain management in organizations.

5.3.1.5 Forecasting

Every company needs to do at least some forecasting in order to strategize and plan; the future success of any business depends heavily on the ability of its management to forecast well.[2] However, the availability of 'good' data is crucial for the use of forecasting methods; otherwise, it would turn into 'garbage in, garbage out'. The accuracy of forecasts and the efficiency of subsequent production and service planning are related to the stability and consistency of the processes which are, in turn, influenced by successful applications of standard Six Sigma tools. Six Sigma tools and methods identify and eliminate process defects and diminish process variation. Six Sigma also requires

that data be collected in an accurate and scientific manner. The combination of defect elimination, variation diminishing, and more accurate scientific data collection allows forecasting to be conducted more easily and effectively, which will, in turn, help to improve the effectiveness of production and service planning, operations scheduling and management. On the other hand, if the processes are erratic, then forecasting and subsequent production and service planning and operations scheduling will be much less effective or useful. Important applications of forecasting techniques within the context of operations management include demand forecasting, yield forecasting, and inventory forecasting which is essentially the conjunction of the first two. In addition, forecast results are important inputs to other OR/MS techniques such as mathematical programming, queuing, simulation and modeling.

5.3.2 A roadmap that integrates OR/MS techniques

In the development of the new curriculum, we also consider the deliverables for each of the DMAIC phases. Table 5.3 presents a matrix relating the deliverables and an integrated toolset following the DMAIC roadmap.

The type of training BBs should receive is a function of the environment in which they work,[25] and training curricula should be designed accordingly. It is also important in the presentation of the tools to provide roadmaps and step-by-step procedures for each tool and each overall method.[25] The characteristics of Six Sigma that make it effective are the integration of the tools with the DMAIC improvement process and the linking and sequencing of these tools.[25] While most curricula proposed in the literature manifest their integration,[26,27] the linking and sequencing of the tools is less apparent.[25] In this chapter, leveraging on previous programs and our consulting experience, a sequence of deliverables and the associated tools needed in a typical BB project is conceived, bearing in mind the tasks that need to be accomplished in DMAIC phases and the applicability of traditional Six Sigma techniques together with those techniques outlined in Section 3.1.

Table 5.3 presents a matrix that summarizes the DMAIC framework for both manufacturing/operational and transactional environments. The vertical dimension of the matrix lists the deliverables in each DMAIC phase and the horizontal dimension lists the tools/techniques that could be used to serve the purposes in the vertical dimension. The flow of deliverables is self-explanatory as they represent tasks/milestones in a typical DMAIC process. The toolset across the horizontal dimension has been fortified with OR/MS techniques to meet the higher expectation of Six Sigma programs in delivering value to an enterprise. It should be noted that while it is conceivable that a specific OR/MS technique could be applied in multiple phases, we have made each basic OR/MS technique appear with intentional precision within the DMAIC process corresponding to its major areas of application for the purpose of conciseness. Through experience, literature and case study reviews, these are the areas where a majority of Six Sigma projects would benefit from the proposed OR/MS tools. Nonetheless, the placement of various techniques is by no means rigid, due to the broad scope of coverage in Six Sigma projects. The matrix can be used as a roadmap for BBs to implement their projects and as a training curriculum for a new breed of Six Sigma BBs. While more elaborate techniques can also be included, it is felt that the current toolset is the most essential and can be covered in a typical 4-week training program for BBs.

Table 5.3 A Roadmap that Integrates OR/MS Techniques

TRAINING PHASES		DEFINE				MEASURE									ANALYZE																IMPROVE									CONTROL				
IMPLEMENTATION PHASES	TOOLS AND TECHNIQUES / DELIVERABLES	Definition, leadership and implementation issues	Project selection	Probabilistic risk thinking and strategic planning	Decision analysis	Project management tools	Quality function deployment and Kano analysis	Gap analysis	Process mapping	Data sampling techniques (data quality and quantity)	Measurement systems analysis	SPC Part I (concepts, implications of instability)	Run charts	Monte Carlo simulations and statistical distributions	Capability analysis	Exploratory data analysis	Failure modes and effects analysis	Cause–effect analysis	Statistical hypothesis testing	Confidence intervals	Analysis of variance (ANOVA)	Correlation and simple linear regression	Reliability models and measures	Cost analysis	Forecasting	Basic queuing systems	Simulations and statistical distributions	Design of experiments	Response surface methodology	Advanced experimentation: robust design concepts and techniques	Optimization and control of queues	Mathematical programming techniques	Heuristics	Sensitivity analysis	Mistake proofing	Validation testing	Control plans	SPC II: control charts	Basic control charts
	Understanding Six Sigma	•																																								
DEFINE	Formulating strategies with emphasis on risk			•	•			*																																	
DEFINE	Translating strategies into action plans		•					*																																	
DEFINE	Selecting and scoping of projects		•		•	•		*				□												•							•										
DEFINE	Planning of projects			•		•																																				
MEASURE	Listening to the voice of customers						•																																			
MEASURE	Identifying processes								•																																	
MEASURE	Identifying potential KPIVs and KPOVs						•		*									•																								
MEASURE	Translating customers' needs into business requirements						•																																			
MEASURE	Understanding and learning from data									•		□	•	□		•			•	□	•	•	□																			
MEASURE	Understanding and dealing with uncertainties			•	•					•	•			□					•	□							*														
MEASURE	Understanding instability and variations										•	□	•	□		•					•						*														
MEASURE	Knowing current capabilities							*	*			□			•			•																								
ANALYZE	Analyzing data												•	□		•	•	•	•	□	•	•	□		*	*	*															
ANALYZE	Assessing risks				•						•			□			•		•	□	•		□																		
ANALYZE	Estimating cost components																						□	*	*																
ANALYZE	Performing scenario/what if analysis																		•	□	•	•			*		•					•									
ANALYZE	Analyzing job sequences and cycle time																									*	*	•													
ANALYZE	Evaluating options and work-flow designs				•				*								•	•	•	□	•	•	□	*		*	*	•				*									
IMPROVE	Selecting KPIVs and KPOVs																				•	•	□					•				*	*								
IMPROVE	Establishing transfer function																									*	*	•	□												
IMPROVE	Improving performance measures																					•	□						□		*	*	*	•							
IMPROVE	Reducing variability in processes																												□	□	*	*	*	•							
CONTROL	Implementing robust processes																																	•	•		•				
CONTROL	Verifying achievement													•					•		•						•									•		□	*		
CONTROL	Sustaining Improvements																																			•		□	*		

Six Sigma is well known to be a highly applied and result-oriented quality engineering framework and curriculum as compared to other programs such as the Certified Quality Engineer programs run by the American Society of Quality.[26] The basis of its strength does not lie in each individual tool but in the effective integration of the various tools, with a strong emphasis on statistical thinking in the reduction of variability in products and processes. However, as discussed in the preceding sections, tools and techniques within the existing Six Sigma framework are inadequate to deal with many problems in product and service delivery processes. In a bid to close the gap, a stronger Six Sigma toolset containing OR/MS techniques capable of dealing with many of such problems has been proposed. The linking and sequencing of the proposed tools are driven towards a practical integration within the Six Sigma DMAIC framework.

5.4 CASE STUDY: MANPOWER RESOURCE PLANNING

This case study focuses on the application of the newly proposed OR/MS tools for efficient manpower resource allocation. Some details associated with the Six Sigma project have been deliberately left out for reasons of confidentiality. However, pertinent infomation relating to the analysis with the newly proposed OR/MS tool is provided to demonstrate its effectiveness.

Here, we give an illustration of how the Six Sigma framework is applied to reduce the waiting times for a retail pharmacy in a hospital. This investigation was prompted by complaints about the long waiting times for drug prescriptions to be dispensed. Initially, continuous efforts were made to expedite the work flow by the pharmacy staff without a systemic examination of relevant work processes. These efforts turned out to be insufficient in achieving the waiting time target set by the management for the pharmacy.

In the Define phase, external Six Sigma consultants, together with the hospital management, selected this particular project to reduce the waiting times of patients in the central pharmacy given the urgency and proximity of the process to the customers. Furthermore, as there were multiple satellite pharmacies with similar processes to those of the central pharmacy, the project would be able to reap benefits beyond this particular department. The success of this project would enable the Six Sigma team to garner more extensive buy-in and support from the management and other hospital staff. Such internal 'marketing' efforts are essential for sustainable implementation and successful execution of future Six Sigma projects that will deal with increasingly difficult problems.

Four major tasks were identified in the current process – *typing*, *packing*, *checking*, and *checking and dispensing* (or *dispensing* for short). The arrival rates of prescriptions to the pharmacy were measured based on counts of arrivals each 10 minutes. The profile of estimated arrival rates at each 10 minutes interval is shown in Figure 5.1. After accounting for outliers in the profile and reasons for high arrival rates in some particular instances as shown in Figure 5.1, the profile was discretized and two distinct arrival rates identified by visually examining the data. Although the profile can be more accurately discretized by having additional segments, only two distinct arrival rates were identified in preliminary investigations after accounting for practical considerations related to manpower allocations and sources of variations,

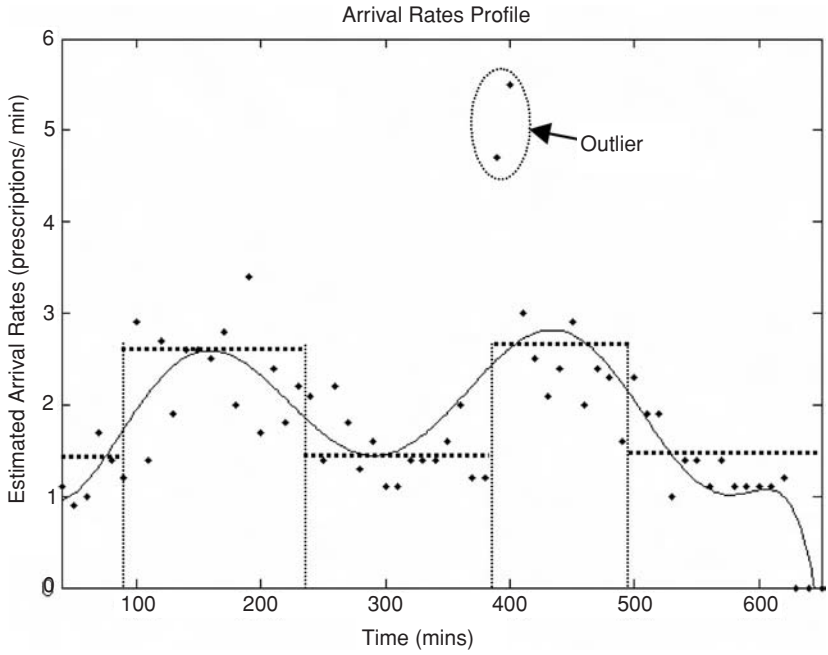


Figure 5.1 Profile of estimated arrival rates.

such as day-to-day variations, in the arrival rates. Furthermore, subsequent sensitivity analysis based on different arrival rates using queuing methodologies will allow us to select the most robust manpower configurations for each distinct arrival rate.

With reasonable estimates of data on the arrival rates extracted from the arrival rate profile shown in Figure 5.1, together with service and rework rates shown in Table 5.4 obtained in the Measure phase, the lead times and value-added times for each process can be computed. The entire process can in fact be represented by an open queuing network as shown in Figure 5.2.

Table 5.4 Service and rework rates.

Process	Estimated service rates (min/job)
Typing	1.92
Packing	0.20
Checking	1.70
Checking and dispensing	0.19
Rework routing	**Estimated proprotion of rework**
Packing→Typing	0.025
Checking→Typing	0.025
Checking→Packing	0.025
Checking and Dispensing→Typing	0.001
Checking and Dispensing→Packing	0.001

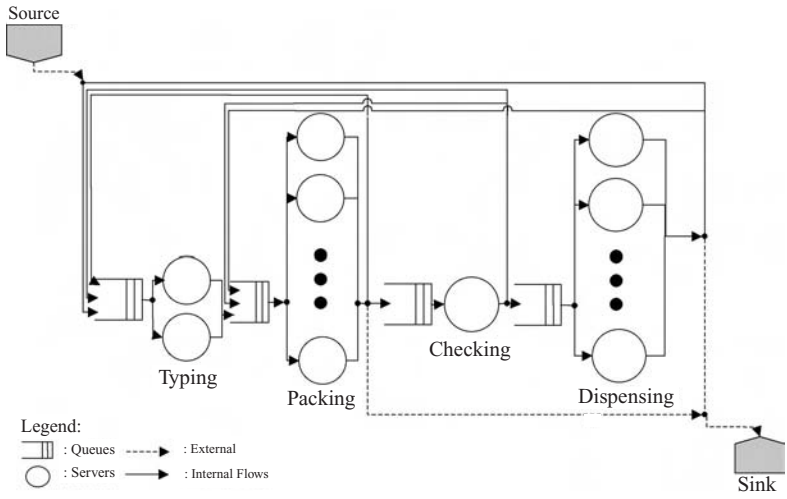


Figure 5.2 Queuing network representation of drug dispensing process.

For the Analyze phase, estimates of average waiting times and queue lengths can be derived using basic queuing methodologies based on the patient arrival rates and service rates. Steady-state queuing analysis is adequate here as the arrival and service rates are fast enough to ensure the system reaches its steady state in a short time. Inter-arrival and service times were assumed to be exponentially distributed. The entire process can thus be represented as a system of interconnected $M/M/1$ and $M/M/s$ service stations. $M/M/1$ and $M/M/s$ are standard abbreviations to characterize service stations in queuing methodologies, $M/M/1$ denoting service stations with a single server, and $M/M/s$ stations with a finite number s of servers ($s > 1$). Both of these types of stations experience Markovian arrival and service processes, with service processes in an $M/M/s$ system being independently and identically distributed. Furthermore, the queuing buffer is assumed to be of infinite size and each server can only serve one customer at a time and selects waiting customers on a first come, first served basis.

Given the preceding assumptions, mean total waiting times for the entire drug dispensing process in the pharmacy can be computed by first computing the mean sojourn times using standard queuing formulas for each service station,[28] and then summing these mean sojourn times. In the computations, the mean total waiting time for the entire process does not include the mean service time of the final dispensing and checking process. This is because the waiting time of a patient is defined as the period from the time he submits the prescription to the time when service by the dispensing pharmacist is initiated.

5.4.1 Sensitivity analysis

At the Improve phase, the impact of different manpower configurations on the overall waiting times can be assessed by varying the number of packers and dispensing

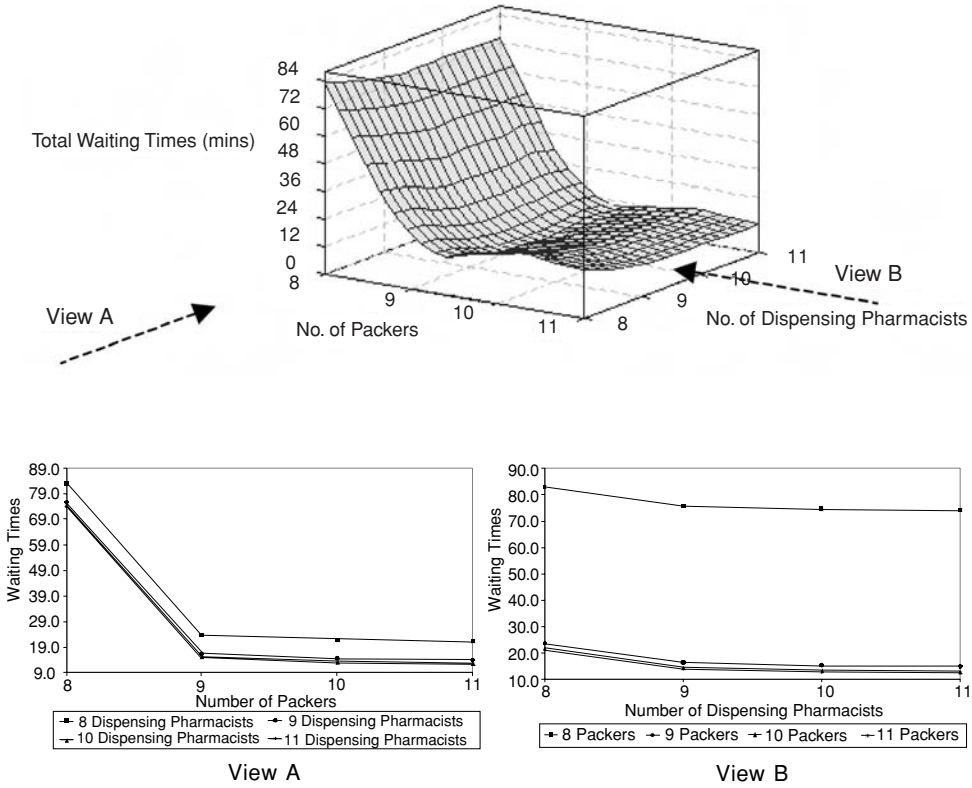


Figure 5.3 Sensitivity of total waiting times to variations in the numbers of packers and dispensing pharmacists.

pharmacists in a sensitivity analysis. Results for the lower arrival rate extracted from the estimated arrival rate profile shown in Figure 5.1 are presented here. A corresponding analysis for the higher arrival rate can be conducted. Subject to other practical considerations, manpower deployment can then be adjusted dynamically throughout the day.

With an arrival rate of 88 prescriptions an hour, at least 8 packers and 8 dispensing pharmacists were needed for the packing and checking dispensing processes, respectively, in order to ensure the finiteness of steady-state waiting times. Figure 5.3 shows the impact of varying the number of packers and dispensing pharmacists on the mean total waiting times. It was observed that waiting times would be increased significantly if the number of packers was reduced to 8. With more than 8 packers, the waiting times were observed to be relatively stable over the different numbers of dispensing pharmacists in the experiment. It was further observed that by having one additional dispensing pharmacist, the targeted 15 minutes of mean total waiting times could potentially be met.

In order to assess the robustness of each possible system configuration, arrival rates of patients at the pharmacy were varied in the model. Figure 5.4 shows the

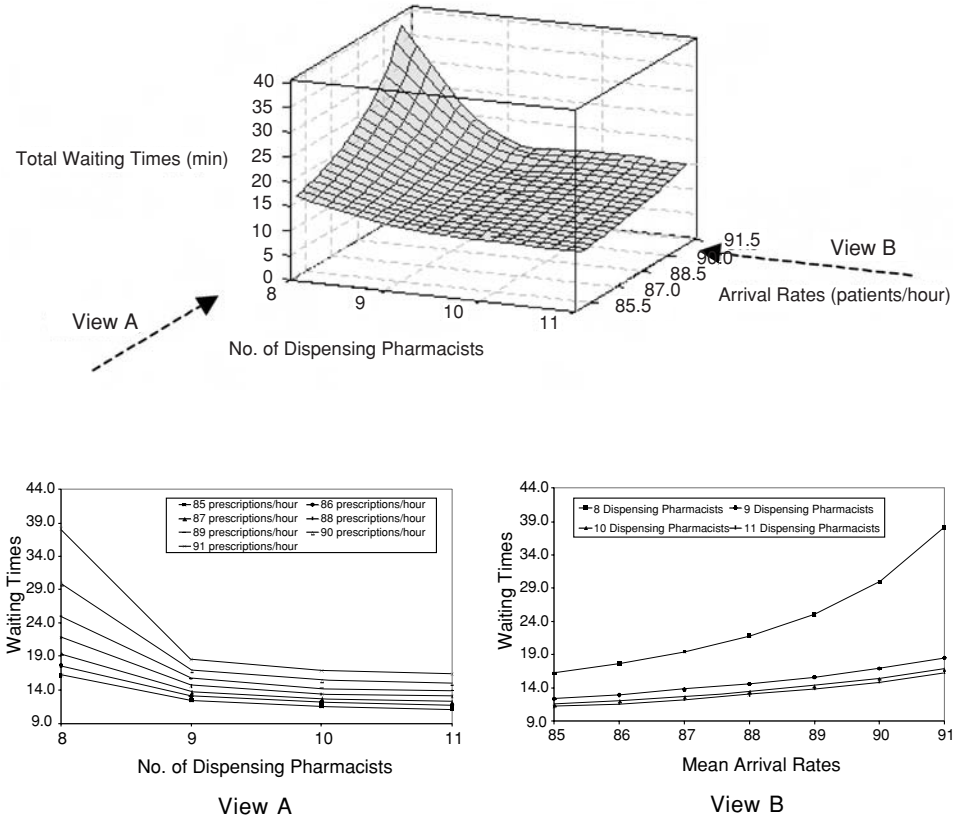


Figure 5.4 Sensitivity of total waiting times to variations in number dispensing pharmacists and patients arrival rates.

sensitivity of mean total waiting times subjected to small perturbations in arrival rates for different numbers of dispensing pharmacists. It was observed that the original configuration with only 8 dispensing pharmacists would result in large increases in waiting times when there were only small increases in the number of prescriptions arriving per hour (in the range 85–91 prescriptions per hour). This provided the management with insights into the frequently experienced phenomenon of doubling in waiting times on some 'bad' days. By having an additional dispensing pharmacist, the system would be expected to experience a more stable waiting time distribution for different arrival rates.

In order to elicit more improvement opportunities, additional root cause analysis for the long waiting times was performed for the present process. This was conducted with the aid of a fishbone (or Ishikawa) diagram, enabling. The potential of streamlining manpower deployment in the pharmacy can be further studied. From the analysis, it was suggested that a new process, that screening, be implemented at the point when the pharmacy receives prescriptions from patients (prior to the existing typing

Table 5.5 Comparisons of sojourn times for each process and mean total waiting times.

Type of job	Without screening	With screening
Screening	–	2.6 (1 pharmacist)
Typing	0.6 (2 typists)	0.6 (2 typists)
Packing	5.9 (10 packers)	5.6 (10 packers)
Checking	6.2 (1 pharmacist)	5.0 (1 pharmacist)
Dispensing checking*	9.0 (8 pharmacists)	2.0 (7 pharmacists)
Mean total waiting time	21.7	15.8

*Service times only

process). It was expected that this would help to alleviate problems associated with errors in prescriptions and medicine shortages.

A pilot run was implemented with one dispensing pharmacist moved to the newly implemented screening process. Such a reconfiguration would not require additional pharmacists to be hired. Service and arrival rates estimates were obtained from the pilot runs and mean waiting times predicted for each process. The mean total waiting times for prescriptions can again be computed by summing all the mean waiting and service times of each process. The average waiting times computed were statistically validated with actual data obtained from pilot runs.

Table 5.5 shows the improvements in the average sojourn times of each subprocess and the mean total waiting time for processes before and after the addition of the screening process. Improvements can be observed in the new process because many interruptions that occurred during the dispensing process were effectively reduced by the screening process upfront. As a result, the productive time of pharmacists increased and the mean queue length in front of the dispensing process shortened from 13 to 3.

Various possible system configurations were again tried with different number of packers and dispensing pharmacists with the new model that considered the screening process. In order to ensure finiteness of steady-state waiting times, at least 8 packers and 7 dispensing pharmacists for the packing and "checking and dispensing processes respectively are needed. From the analysis, the proposed new configuration was found to be more robust to changes in manpower deployment over the packing and dispensing subprocesses (see Figure 5.5). Eventually, this new robust design was adopted to ensure waiting-time stability over possible variations in manpower deployment.

Process validations on results generated from the mathematical models depicting original and improved processes have been dealt with throughout the Analyze and Improve phases in the preceding discussions. At the final Control phase, standard operating procedures (SOPs) were put in place in order to ensure stability the of new processes. During the generation of these SOPs, several new control measures were proposed by the team and implemented with the understanding and inputs from the

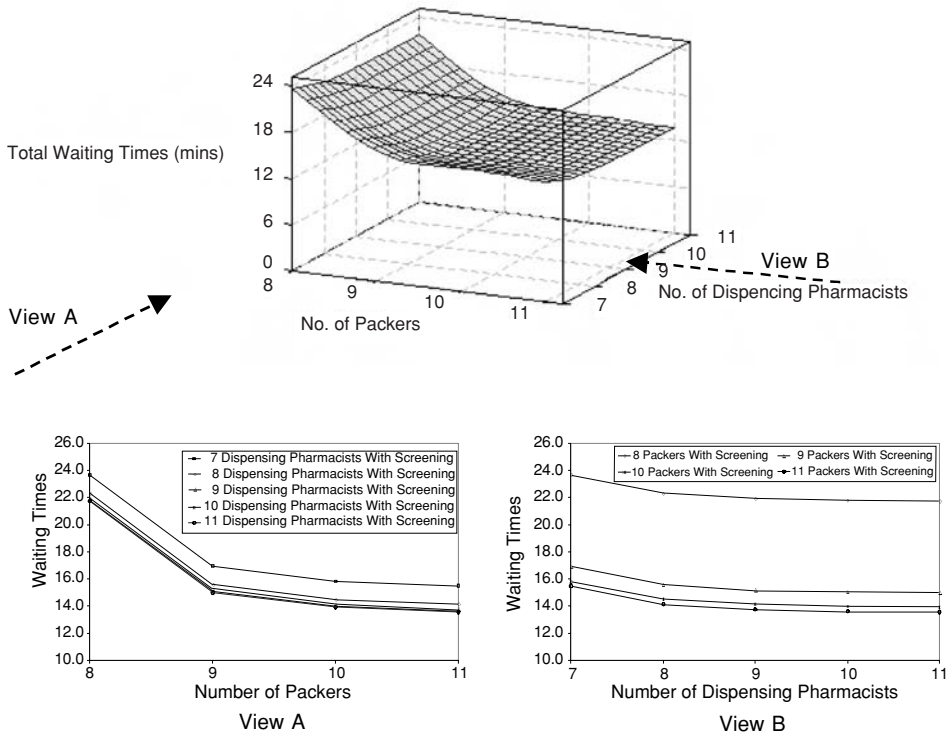


Figure 5.5 Sensitivity of total waiting times to variations in number of packers and dispensing pharmacists in the new dispensing process with screening.

relevant stakeholders through targeted 'Kaizen' events. The SOP and validated results were communicated to the pharmacy staff in order to reinforce their confidence in the new processes.

5.4.2 Relaxing exponential assumptions on service times distribution

Further sensitivity analysis can be performed to assess the effect of the exponential assumptions on the distribution of service times. This analysis was conducted to establish whether the assumption of exponentially distributed service times would result in more conservative system design choices. Results based on the model without the screening process are described here. A similar sensitivity analysis can be conducted on the model with the new screening process.

Each service station i was assumed to experience a Poisson arrival process with mean arrival rate λ_i. Single-server service stations i were assumed to experience service processes that follow any general distribution with mean service rate μ_i. This is commonly known as the $M/G/1$ queuing system. For service stations with (finitely) many servers, the service processes of each server were assumed to follow general

distributions that were independently and identically distributed with mean service rate μ_i. Such a system is commonly known as the $M/G/s$ queuing system ($s > 1$ finite). The mean total waiting time for the entire drug dispensing process in the pharmacy can be computed by summing the mean waiting and mean service times at each service station.

In order to derive the mean waiting times at each service station, the overall arrival rate λ_i at each service station, for the queuing network shown in Figure 5.2 has to be computed. At statistical equilibrium, λ_i is given by

$$\lambda_i = \lambda_{i0} + \sum_{j=1}^{N} \lambda_j P_{ji},$$

where N is the number of queuing stations in the network, λ_i is the arrival rate at station i from external sources, and p_{ji} is the probability that a job is transferred to the jth node after service is completed at the ith node.

In order to compute the mean waiting times at each single-server queuing stations ($s_i = 1$), the mean queue length of each single server-station, $\bar{L}_i^{M/G/1}$, is first computed with the well-known *Pollaczek–Khintchine* formula,[28]

$$\bar{L}_i^{M/G/1} = \frac{\rho_i^2}{1-\rho_i} \frac{1 + C V_i^2}{2}.$$

Here $C V_i^2$ is the squared cofficient of variation of service times, T_i, which follow any general random distribution:

$$C V_i^2 = \frac{\text{Var}(T_i)}{\bar{T}_i^2},$$

In which $\text{Var}(T_i)$ is the variance of the random service time of server i, and $\bar{T}_i$ is the mean of the service time of server i (or reciprocal of the service rate μ_i). ρ_i is the utilization of server i, which can be interpreted as the fracttion of time during which the server is busy and is given by

$$\rho_i = \frac{\lambda_i}{\mu_i}.$$

Given $\bar{L}_i^{M/G/1}$, the mean waiting times at each single-server service station, $\bar{W}_i^{M/G/1}$, can then be computed using *Little's theorem*[28] as follows:

$$\bar{W}_i^{M/G/1} = \frac{\bar{L}_i^{M/G/1}}{\lambda_i}.$$

To compute the waiting times of queuing stations with multiple server ($s_i > 1$), we apply an approximation due Cosmetatos.[29] For this, we first compute the mean

waiting times of an $M/M/s$ queuing system ($\bar{W}_i^{M/M/s}$):

$$\bar{W}_i^{M/M/s} = \frac{\bar{L}_i^{M/M/s}}{\lambda_i}$$

where the mean queue length of the $M/M/s$ system, $\bar{L}_i^{M/M/s}$, is given by:

$$\bar{L}_i^{M/M/s} = \frac{(s_i\rho_i)^{s_i}\rho_i P_{i0}}{s_i!(1-\rho_i)^2}.$$

In this expression P_{i0} is the probability of station i begin empty,

$$P_{i0} = \left[\sum_{n=0}^{s_i-1}\frac{(s_i\rho_i)^n}{n!} + \frac{(s_i\rho_i)^{s_i}}{s_i!(1-\rho_i)}\right]^{-1},$$

and ρ_i is the utilization of service station i, this time multiple servers,

$$\rho_i = \frac{\lambda_i}{s_i\mu_i}.$$

Next, we compute the mean waiting times of an $M/D/s$ queuing system, that is, for s servers with constant (i.e deterministic) service times ($\bar{W}_i^{M/D/s}$), as follows:

$$\bar{W}_i^{M/D/s} = \frac{1}{2}\frac{1}{K_i}\bar{W}_i^{M/M/s},$$

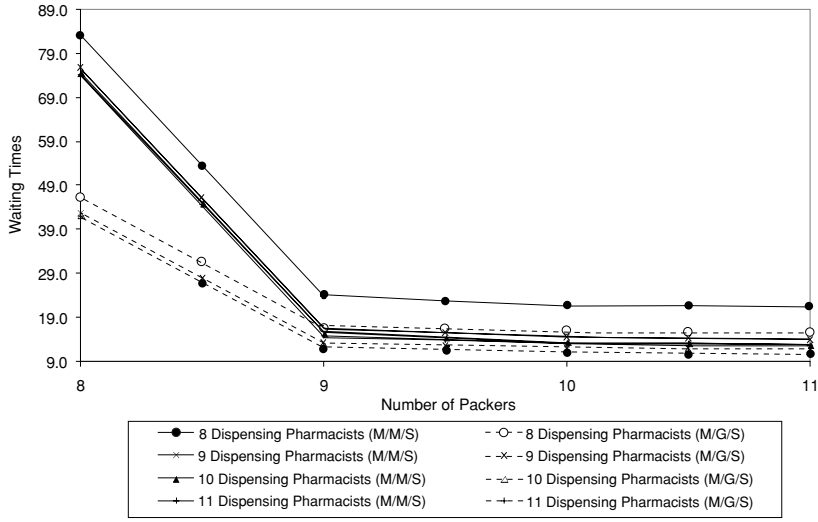
where

$$K_i = \left(1 + (1-\rho_i)(s_i-1)\frac{\sqrt{4+5s_i}-2}{16\rho_i s_i}\right)^{-1}.$$

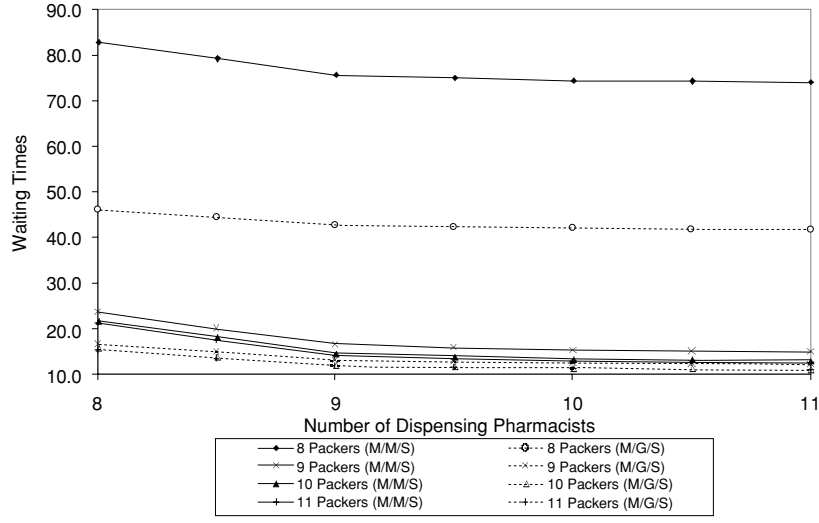
The approximate mean waiting times for an $M/G/s$ system can finally be computed from:

$$\bar{W}_i^{M/G/s} \approx C\,V_i^2\,\bar{W}_i^{M/M/s} + \left(1 - C\,V_i^2\right)\bar{W}_i^{M/D/s}.$$

Figure 5.6 shows the difference in mean waiting times computed with and without exponential service times assumptions. It was observed that the mean total waiting time was higher when service times were assumed to be exponentially distributed than to when no such assumption was made. In many service processes, the exponential assumption of the distribution of service times usually results in more conservative queuing system designs. This is because, given the same system configurations, the



(a)



(b)

Figure 5.6 Comparisons of mean total waiting times computed with and without assumptions of exponential service times.

expected waiting times predicted with queuing models assuming exponentially distributed inter-arrival and service times will be higher than with models assuming any other distributional assumptions whose coefficient of variation is less than unity. Decisions based on mean waiting times and queue lengths predicted from such models would thus err on the safe side.

5.5 CONCLUSIONS

Regardless of which industrial sector a BB is being employed in, he needs to adopt a systems view of the operations of an enterprise. Current BB training programs are no longer adequate for the increasingly demanding customers of the twenty-frist century. A new breed of BBs will need to integrate OR/MS techniques into their Six Sigma toolset so that it can remain relevant. A new roadmap is formulated and presented in Table 5.3 to meet these emerging needs.

Not all the OR/MS tools will be used in a project, but they serve as a reminder/checklist. In this way, a BB can remain focused on the project while being alert to other tools that may be useful in providing a solution. It could be argued that a Six Sigma BB armed with OR/MS techniques would operate like a 'Super Belt', with breath and depth well beyond what is found in the routine toolset of BBs coming from a regular Six Sigma training conveyor belt.

In addition to OR/MS techniques, there is also an emerging trend of integrating artificial intelligence and information systems technologies, such as data mining,[8] fuzzy logic and neural networks, into Six Sigma programs – in particular, DFSS for software development. As the scope of Six Sigma application expands with time, more cross-functional tools will be integrated with Six Sigma to achieve even wider and deeper business performance improvement. The current integration of OR/MS tools is only part of the itinerary in the journey towards Six Sigma excellence.

REFERENCES

1. Goh, T.N. (2002) A strategic assessment of Six Sigma. *Quality and Reliability Engineering International*, 18, 403–410. See also Chapter 2, this volume.
2. Hillier, F.S. and Lieberman, G.J. (2001) *Introduction to Operations Research*, 7th edition Boston: McGraw-Hill.
3. Harry, M.J. and Schroeder, R. (2000) *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations*. New York: Dowbleday.
4. Snee, R.D. (2000) Guest editorial: Impact of Six Sigma on quality engineering. *Quality Engineering*, 12(3), ix–xiv.
5. Montgomery, D.C. (2001) Editorial: Beyond Six Sigma. *Quality and Reliability Engineering International*, 17(4), iii–iv.
6. Tang, L.C. and Paoli, P.M. (2004) A spreadsheet-based multiple criteria optimization framework for Quality Function Deployment. *International Journal of Quality and Reliability Management*, 21(3), 329–347.
7. Recker, R. and Bolstorff, P. (2003) *Integration of SCOR with Lean & Six Sigma*. Supply-Chain Council, Advanced Integrated Technologies Group.
8. Goh T.N. (2002) The role of statistical design of experiments in Six Sigma: Perspectives of a practitioner. *Quality Engineering*, 14(4), 659–671.
9. Martello, S. and Toth, P. (1990) *Knapsack Problems: Algorithm and Computer Implementations*. New York: John Wiley & Sons, Inc.
10. Zhang, C.W. and Ong, H.L. (2004) Solving the biobjective zero-one knapsack problem by an efficient LP-based heuristic. *European Journal of Operational Research*, 159(3), 545–557.
11. Tang, L.C. and Xu, K. (2002) A unified approach for dual response surface optimization. *Journal of Quality Technology*, 34(4), 437–447. See also Chapter 20, this volume.
12. Lam, S.W. and Tang, L.C. (2005) A graphical approach to the dual response robust design problem. *In Proceedings of the Annual Reliability and Maintainability Symposium*, pp. 200–206. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

13. Harry, M. and Stewart, R. (1988) *Six Sigma Mechanical Design Tolerancing*. Schaumbwg, IL: Motorola University Press.
14. Harry, M.J. and Lawson, J.R. (1992) *Six Sigma Producibility Analysis and Process Characterization*. Reading, MA: Addison-Wesley.
15. Tagaras, G. (1989) Power approximations in the economic design of control chart. *Naval Research Logistics*, 36, 639–654.
16. Crowder, S.V. (1992) An SPC model for short production runs: minimizing expected cost. *Technometrics*, 34(1), 64–73.
17. Rahim, M.A. (1993) Economic design of $\bar{X}$ control charts assuming Weibull in-control times. *Journal of Quality Technology*, 25(4), 296–305.
18. Chung, K.J. (1994) Algorithm for computing the economically optimal $\bar{X}$ control chart for a process with multiple assignable causes. *European Journal of Operational Research*, 72(2), 350–363.
19. McWilliams, T.P., Saniga, E.M. and Davis, D.J. (2001) Economic-statistical design of $\bar{X}$ and R or $\bar{X}$ and S charts. *Journal of Quality Technology*, 33(2), 234–241,.
20. Rohleder T.R. and Silver E.A. (2002) Statistical process control subject to a labor resource constraint. *International Journal of Production Research*, 40(14), 3337–3356.
21. Hahn, G.J., Doganaksoy, N. and Hoerl, R. (2000) The evolution of Six Sigma. *Quality Engineering*, 12(3), 317–326.
22. Bayle, P., Farrington, M., Sharp, B., Hild, C. and Sanders, D. (2001) Illustration of Six Sigma assistance on a design project. *Quality Engineering*, 13(3), 341–348.
23. Taha, H.A. (2003) *Operations Research: An Introduction*, 7th edn. Upper Saddle River, NJ: Pearson Education International.
24. Prabhu, N.U. (1965) *Queues and Inventories: A Study of Their Basic Stochastic Processes*. New York John Wiley & Sons, Inc.
25. Snee, R.D. (2001) Discussion. *Journal of Quality Technology*, 33(4), 414–417.
26. Montgomery, D.C., Lawson, C., Molnau, W.E. and Elias, R. (2001) Discussion. *Journal of Quality Technology*, 33(4), 407–409.
27. Hahn, G.J., Hill, W.J., Hoerl, R.W. and Zinkgraf, S.A. (1999) The impact of Six Sigma improvement – A glimpse into the future of statistics. *American Statistician*, 53, 1–8.
28. Little J. (1961) A proof of a queuing formula. *Operations Research*, 9(3), 383–387.
29. Cosmetatos, G.P. (1976) Some approximate equilibrium results for multi-server queues $(M/G/r)$. *Operational Research Quarterly*, 27(3), 615–620.

Part B

Measure Phase

6

Process Variations and Their Estimates

L. C. Tang and H. S. Yam

The key to successful statistical analysis is the identification of variations, their sources and their estimates. Variation is inherent in almost everything. In order to achieve a given business target, the capabilities and variability of the process must be well established and controlled. In this chapter the sources of variation will be discussed, followed by the basic statistical properties of the process variability σ^2 and its unbiased estimator. A short review of the nested design with variance components is then presented to illustrate how various sources of variation can be estimated.

6.1 INTRODUCTION

Business growth depends on how well companies meet customer expectations in terms of quality, price and delivery. Their ability to satisfy these needs with a known degree of certainty is controlled by process capability, and the amount of variation in their processes. Variation has a direct impact on business results in terms of cost, cycle time, and the number of defects that affect customer satisfaction. No two units of product produced by a manufacturing process (or two services provided by a service provider) are identical. Variation is inherent in all processes and measurements, natural or human-made. Figure 6.1 shows two main sources of variation, which are *process variation* and *measurement variation*.

6.1.1 Process variation

Process variation is important in the Six Sigma methodology, because the customer is always evaluating the services, products and processes provided to them, to determine

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

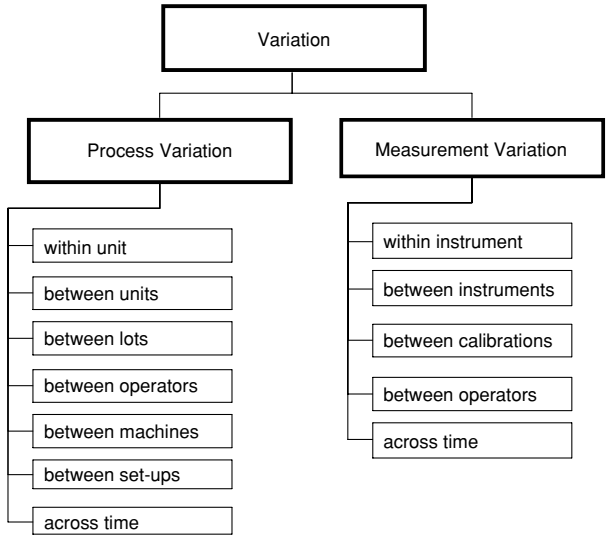


Figure 6.1 Sources of variation.

how well these services, products and processes are meeting their requirements and expectations; in other words, how well they conform to standards. Distinguishing the sources of process variation is paramount because the appropriate measurements are different for each. Without this distinction, five-step problem-solving approach of the Six Sigma methodology – define, measure, analyze, improve, and control (DMAIC) cannot be utilized effectively (the detailed and extended DMAIC approach can be found in the book by Harry and Schroeder).[1]

Many processes have more than one source of variation in them. In order to reduce variation in processes, these multiple sources must be understood. As shown in Figure 6.1, process variation consists of within-unit variation, between-units variation, between-lots variation, between-operators variation, between-machines variation, between-set-ups variation, and, lastly, variation over time.

For example, in the semiconductor industry, a batch process may operate on several wafers at a time. The defects found at different locations on a wafer are considered as a within-unit variation, as there are many other locations where the wafer is defect-free. On the other hand, one might detect the defect at one particular location of one wafer but not of another wafer in the same process batch; such variation is categorized as between-unit variation. The difference in defective rate between each lot is considered as between-lots variation. Similarly, the disparity between the products or services provided by different operators or machines is designated as between-operators variation, or between-machines variation, respectively. The difference in process yield for different set-ups can be treated as between-set-ups variation. Due to aging, the performance of the process will deteriorate, and this variation is unavoidable over the lifetime of a process. In addition, the cumulative effect of many small, essentially unavoidable causes will lead to different process outcomes; these chance causes are an inherent part of the process.

6.1.2 Measurement variation

Another main source of variation is measurement variation. In continuous improvement programs, design and appropriateness qualities gain their significance only after measurement activities.

The quality of measurement data depends on the statistical properties of multiple measurements. There are many factors leading to the variability of the measured quality characteristic against the real value. If the instrument fails to repeat the same value of measurement, this is defined as within-instrument variation and is called "repeatability" in measurement systems analysis (MSA).[2] On the other hand, variation among the measurement instruments occurs when different instruments give different readings for the same quality characteristic of a specimen. The measuring device has to be calibrated from time to time. Different calibration methods will result in different degrees of precision of the measuring device and thus lead to the variation between calibrations. The variability due to different operators using the instrument is between-operators variation and is called "reproducibility" in MSA. Also, there will be unavoidable variability across time.

6.1.3 The observed variation

Generally, in any activity involving measurements, some of the observed variability will be inherent in the units or items that are being measured, and some of it will result from the measurement that is used. The variance in the recorded observations, σ^2_{tot}, is comprised of two components: the variance of the product (process characteristic), σ^2_{proc} and the variability of the measurement system σ^2_{meas}:

$$\sigma^2_{\text{tot}} = \sigma^2_{\text{proc}} + \sigma^2_{\text{meas}}.$$

A possible impediment to the effect implementation of the DMAIC approach in problem-solving is measurement error. Consequently, it is important that measurement systems be in a state of statistical control. Figure 6.2 shows the relationship of the two components with the variance of the recorded observations.

With statistical procedures in general and control charts in particular, a primary objective is to analyze components of variability so that variability due to assignable causes can be detected, and to maintain the process with only natural variability, often referred to as a 'stable system with chance causes'. The chance causes of variation

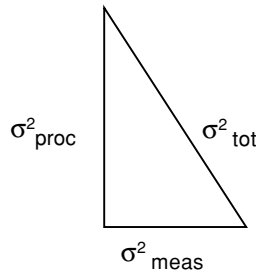


Figure 6.2 Relationship between total σ process σ, and measurement system σ.

are defined as the cumulative effect of many small, essentially unavoidable causes, whereas the assignable causes are generally large compared to the chance causes, and usually represent an unacceptable level of process performance.

It is obvious that the variability of the measurement system used should be as small as possible, so that the variability observed is mainly from the process and so that action can be taken accordingly to reduce process variability.

6.1.4 Sampling and rational subgroups

Distinguishing between the types of variation is paramount because the appropriate managerial actions are different for each. Without this distinction, management will never be able to identify the key elements for improvement. Sampling consists of selecting some part of a population to observe so that one may have some idea about the whole population. Due to time and/or cost constraints, sampling inspection is preferred. From the previous section, it is clear that there are various sources of variation. In order to measure (or control) certain quality characteristics more effectively, the concept of rational subgroups can be employed.

The rational subgroup concept means that subgroups or samples should be selected so that if there is a change in population parameter, the chance of differences between subgroups will be maximized, while the chance of differences due to the change in parameter within a subgroup is minimized. For example, if some of the observations in the sample are taken at the end of one shift and the remaining observations are taken at the start of the next shift, then any differences between shifts might not be detected. The undetected differences between shifts might consist of variation between operators, between machines, and also between set-ups. Thus, if one were to monitor the performances of certain machines, the rational subgroups should be constructed accordingly.

A basic requirement for the applicability of a subgroup approach is that the observations within the subgroup naturally represent some appropriately defined unit of time. In addition, subgroups should be chosen in such a way that the observations within the sample have been, in all likelihood, measured under the same process conditions. By doing so, the subgroup elements will be as homogeneous as possible and thus special causes are more likely reflected in greater variability of the subgroups themselves.

6.2 PROCESS VARIABILITY

The process variability can be measured by the population variance, σ^2. When the population is finite and consist of N values, the population variance is given by

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N},$$

where μ is the process mean. The positive square root of σ^2, or σ, is the population standard deviation. The standard deviation was first proposed by Pearson as a measurement of variation or scatter within a data set.[3]

It is practically impossible to determine the true value of the population parameters μ and σ via a finite sample of size n. Thus, sample statistics are used. Suppose that x_1, $x_2, \ldots, x_n$ are the observations in a sample. Then the variability of the process sample data is measured by the sample variance,

$$s_n^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n},$$

where $\bar{x}$ is the sample mean given by $(\sum_{i=1}^{n} x_i)/n$. Note that the sample variance is simply the sum of the squared deviations of each observation from the sample mean, divided by the sample size. However, the sample variance defined is not an unbiased estimator of the population variance σ^2. In order to obtain an unbiased estimator for σ^2, it is necessary instead to define a 'bias-corrected sample variance',

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}.$$

An intuitive way to see why s_n^2 gives a biased estimator of the population variance is that the true value of the population mean, μ, is almost never known, and so the sum of the squared deviations about the sample mean $\bar{x}$ must be used instead. However, the observations x_i tend to be closer to their sample mean than to the population mean. Therefore, to compensate for this, $n-1$ is used as the divisor rather than n. If n is used as the divisor in the sample variance, we would obtain a measure of variability that is, on the average, consistently smaller than the true population variance σ^2. Another way to think about this is to consider the sample variance s^2 as being based on $n-1$ degrees of freedom since the sample mean is used instead.

If the individual observations are from the normal distribution, the sample-to-sample randomness of s^2 is explained through the following random variable:

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma^2}. \tag{6.1}$$

The random variable χ^2 follows what is known as the *chi-square* distribution[4] with $n-1$ degrees of freedom which is also its mean. Even though the derivation of this statistic is based on the normality of the x variable, the results will hold approximately as long as the departure from normality is not too severe.

6.2.1 The unbiased estimator

The sample mean $\bar{x}$ and the biased-corrected sample variance s^2 are unbiased estimators of the population mean and variance μ and σ^2, respectively. That is,

$$E\bar{x} = \mu \quad \text{and} \quad E(s^2) = \sigma^2.$$

If there is no variability in the sample, then each sample observation $x_1 = \bar{x}$, and the sample variance $s^2 = 0$. Generally, the larger the sample variance s^2, the greater is the variability in the sample data.

While the sample variance provides an unbiased estimation of the population variance, the positive square root of the sample variance, known as the sample standard deviation and denoted by s, is a biased estimator of the population standard deviation

– that is, s does not provide an accurate estimate of σ. In fact, s will generally provide an underestimate of σ, that is s is smaller than the true value of σ. This is due to the fact that s is based on the sample mean, while σ is based on the population mean. By virtue of the property of the sum of squared deviations, $(x_i - \bar{x})^2$ is minimal when $\bar{x}$ is the sample mean. Consequently, $(x_i - \bar{x})^2 < (x_i - \mu)^2$ when $\bar{x} \neq \mu$.

Taking the square root of both sides of equation (6.1) yields a new random variable,

$$Y = \frac{(n-1)^{1/2}s}{\sigma}. \tag{6.2}$$

It turns out that Y follows a distribution known as the *chi* distribution with $n-1$ degrees of freedom. For the *chi* distribution,[5] it can be shown that

$$E(Y) = \sqrt{2}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Substituting (6.2) into (6.3), and rearranging, we obtain

$$\begin{aligned} E(s) &= \sqrt{\frac{2}{n-1}}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\sigma \\ &= c_4\sigma, \end{aligned}$$

where $\Gamma(\bullet)$ is the gamma function and the unbiased estimator of σ is

$$\hat{\sigma} = \frac{s}{c_4}$$

in which c_4 is given by

$$c_4 = \sqrt{\frac{2}{n-1}}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Table 6.1 gives the values of c_4 for $2 \leq n \leq 25$. For simplicity, for $n > 25$, the value of c_4 can be approximated as

$$c_4 \simeq \frac{4(n-1)}{4n-3}.$$

Table 6.1 Values of c_4 for sample size from 2 to 25.

n	c_4	n	c_4	n	c_4
2	0.7979	10	0.9727	18	0.9854
3	0.8862	11	0.9754	19	0.9862
4	0.9213	12	0.9776	20	0.9869
5	0.9400	13	0.9794	21	0.9876
6	0.9515	14	0.9810	22	0.9882
7	0.9594	15	0.9823	23	0.9887
8	0.9650	16	0.9835	24	0.9892
9	0.9693	17	0.9845	25	0.9896

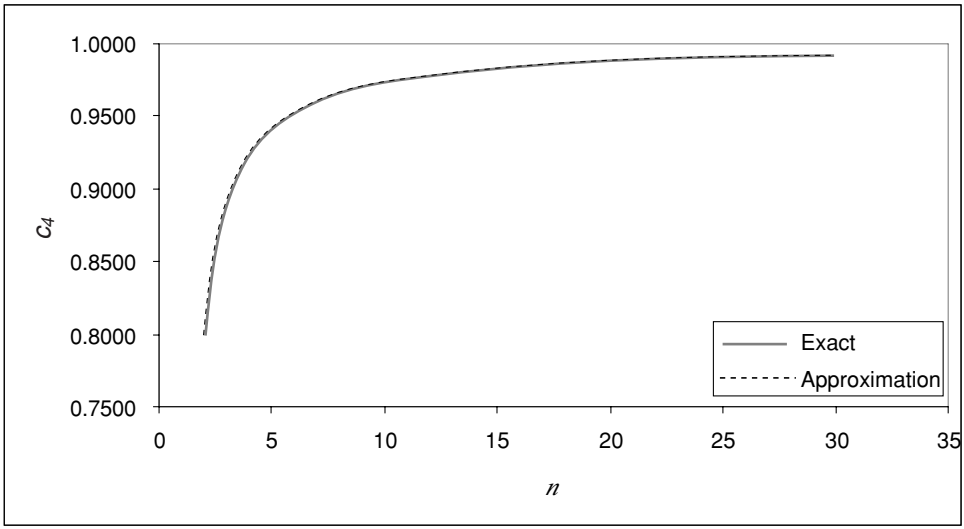


Figure 6.3 Values of c_4 (exact and approximate).

Figure 6.3 shows the c_4 curves for the values obtained from exact calculation as well as the approximation given. From the curves, it is clear that the approximation is closed enough to the values calculated using the exact formula.

6.2.2 A simulation example

When carrying out process characterization and process capability studies, in order to avoid underestimating the process variation and overestimating the process capability, the unbiased estimator s/c_4, instead of the biased estimator s, should be used to estimate the process variation. Otherwise, an incorrect judgment might be made in qualifying a product that is not designed for manufacturability, or in determining the process capability during volume build.

Here, a Monte Carlo simulation is used to show the unbiasedness of the suggested estimator. A normally distributed process output with $\mu = 50$ and $\sigma = 5$ is simulated. Five measurements are taken each day for 30 days. The process is replicated 12 times to simulate the collection of data across a year. The average sample standard deviation (both s_{Bar} and s_{Overall}) over the month is shown in Table 6.2.

It may be observed that correction of s with c_4 provides an unbiased estimate of the known population standard deviation ($\sigma = 5$).

6.3 NESTED DESIGN

A nested design (sometimes referred to as a *hierarchical* design) is used for experiments involving a set of treatments where the experimental units are subsampled, that is, a nested design is one in which random sampling is done from a number of groups

Table 6.2 Simulation results.

Day	x_1	x_2	x_3	x_4	x_5	Mean	StDev	Month	s_{Bar}	$s_{Overall}$	σ_{Bar}	$\sigma_{Overall}$
1	46.23	50.13	50.31	51.96	53.96	50.52	2.85	Jan	5.0559	5.4833	5.3786	5.4925
2	47.16	58.92	52.41	50.56	42.40	50.29	6.15	Feb	4.8114	4.9479	5.1186	4.9593
3	40.96	48.71	53.76	45.53	59.58	49.71	7.23	Mar	4.5384	4.8835	4.8281	4.8917
4	51.00	49.72	47.44	50.68	53.19	50.41	2.09	Apr	4.6134	4.9145	4.9079	4.9227
5	51.18	49.74	51.06	49.94	53.46	51.07	1.48	May	4.6032	4.9449	4.8971	4.9532
6	49.86	51.12	53.61	43.84	47.70	49.22	3.69	Jun	4.5059	4.8724	4.7936	4.8805
7	51.35	49.69	46.00	55.21	43.50	49.15	4.57	Jul	4.7328	5.0265	5.0349	5.0349
8	51.35	59.85	38.18	45.31	54.78	49.89	8.41	Aug	4.6595	4.7991	4.9569	4.8072
9	53.95	54.31	46.84	60.20	46.74	52.41	5.70	Sep	4.7565	5.0178	5.0601	5.0262
10	51.35	61.20	42.96	54.16	56.74	53.28	6.81	Oct	4.8090	4.9671	5.1159	4.9755
11	47.23	47.04	39.94	46.53	45.77	45.30	3.05	Nov	4.5931	4.9188	4.8863	4.9271
12	52.89	36.93	40.94	53.38	45.51	45.93	7.25	Dec	4.5408	4.7698	4.8307	4.7778
13	55.89	57.89	46.04	54.92	53.16	53.58	4.55	Jan	4.8501	5.2811	5.1597	5.2899
14	52.03	46.79	41.74	44.08	46.28	46.18	3.83	Feb	4.7489	5.0618	5.0521	5.0703
15	41.22	49.01	53.25	58.05	54.13	51.13	6.41	Mar	4.8524	5.0646	5.1622	5.0731
16	44.36	38.63	53.24	49.51	45.20	46.19	5.53	Apr	4.0534	4.2794	4.3121	4.2865
17	51.12	44.41	48.26	54.01	34.85	46.53	7.43	May	4.8670	4.9169	5.1777	4.9251
18	45.91	46.72	56.62	43.43	53.33	49.20	5.53	Jun	4.5021	5.0672	4.7895	5.0757
19	52.78	45.60	55.99	50.56	45.82	50.15	4.49	Jul	4.7368	4.9535	5.0392	4.9618
20	49.45	52.89	47.68	57.20	49.02	51.25	3.84	Aug	4.6802	4.8976	4.9789	4.9059
21	60.40	52.55	50.20	51.84	43.97	51.79	5.88	Sep	5.0354	5.5235	5.3568	5.5328
22	49.91	52.06	34.45	48.52	58.47	48.68	8.82	Oct	4.7364	5.0289	5.0388	5.0373
23	46.83	49.20	50.15	52.87	40.27	47.86	4.76	Nov	4.8486	4.9611	5.1580	4.9695
24	46.58	44.28	46.45	43.98	46.36	45.53	1.29	Dec	5.0858	5.0831	5.4104	5.0916
25	42.46	54.59	56.48	53.07	48.69	51.06	5.60	Average	4.4174	4.9860	6.0185	4.9944
26	51.75	58.49	50.26	54.58	55.67	54.15	3.25					
27	52.19	39.34	38.10	42.53	53.89	45.21	7.35					
28	47.29	52.63	36.51	49.14	53.20	47.75	6.75					
29	55.43	54.05	50.26	48.84	43.74	50.46	4.62					
30	55.78	50.81	49.13	51.42	52.27	51.88	2.47					

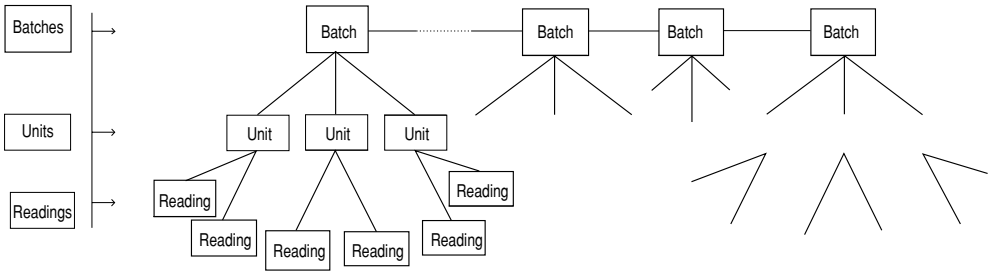


Figure 6.4 A typical nested design.

or batches, and there is no physical relationship among sample 1 from group 1 and sample 1 from group 2. A typical nested design is shown in Figure 6.4.

Suppose, for example, that in the following, two oxide thickness readings from four different wafers produced in 21 different lots are taken (Table 6.3). The average and range (difference between maximum and minimum) for each lot are given in the last two columns. Figure 6.5 gives a graphical representation of this example.

Table 6.3 Oxide thickness for wafers from different lots.

	Wafer 1		Wafer 2		Wafer 3		Wafer 4			
Lot	1	2	1	2	1	2	1	2	Average	Range
1	946	940	966	964	977	946	943	929	951.4	48
2	961	956	977	953	963	977	956	940	960.4	37
3	915	920	948	942	905	887	906	914	917.1	61
4	968	984	970	945	989	962	969	971	969.8	44
5	950	950	962	980	960	954	945	946	955.9	35
6	984	981	967	958	956	960	991	999	974.5	43
7	993	1026	984	959	1034	998	969	992	994.4	75
8	922	935	935	981	927	920	932	939	936.4	61
9	962	968	948	948	933	953	933	921	945.8	47
10	955	945	984	972	946	952	986	984	965.5	41
11	965	957	983	985	982	967	949	933	965.1	52
12	948	928	963	988	943	940	985	986	960.1	60
13	980	970	956	969	1009	1025	970	995	984.3	69
14	964	955	985	988	964	983	963	958	970.0	33
15	950	957	938	925	956	947	934	942	943.6	32
16	952	947	963	978	971	972	975	984	967.8	37
17	949	946	984	993	967	948	1006	1007	975.0	61
18	939	954	933	935	922	921	947	950	937.6	33
19	1010	992	987	987	972	1012	966	963	986.1	49
20	969	1009	958	955	977	979	952	958	969.6	57
21	995	983	999	1002	998	1016	980	975	993.5	41
								Mean	963.0	48.4

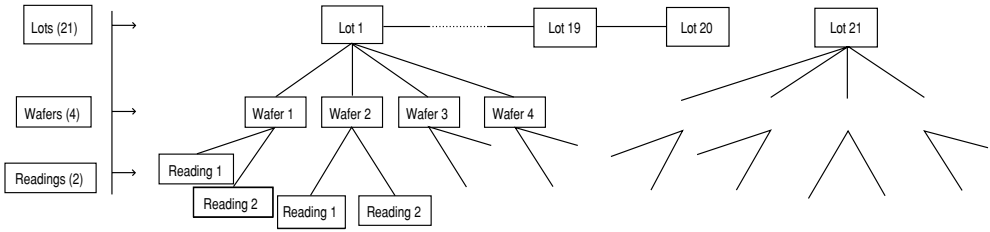


Figure 6.5 Nested design for oxide thickness example.

6.3.1 Variance components and their calculations

Here, an analysis of variance (ANOVA) for the data in Table 6.3 is performed to show the differences between the variance components. From the table, the correlation factor, c, is given by

$$c = N \bullet \bar{Y}^2 = 168(963.04)^2 = 155\,809\,548.21,$$

where N is the total number of readings (measurements) and $\bar{Y}^2$ is the grand mean. The total sum of squares, SS_{tot} is

$$SS_{\text{tot}} = ss_y - c$$

where SS_Y is the sum of square for each measurement. The sum of squares for lots is

$$SS_L = MW \sum_{k=1}^{L} (\bar{y_k} - \bar{y})^2,$$

where L is the total number of lots, M is the total number of measurements in each wafer, and W is the total number of wafers in each lot. Thus, the sum of squares for wafers is

$$SS_W = N \sum_{k=1}^{L} \sum_{j=1}^{W} (\bar{y_{jk}} - \bar{y_k})^2$$

and the sum of squares for measurements is

$$SS_M = \sum_{k=1}^{L} \sum_{j=1}^{W} \sum_{i=1}^{M} (\bar{y_{ijk}} - \bar{y_{jk}})^2;$$

alternatively, SS_M may be obtained by

$$SS_M = SS_{\text{tot}} - SS_L - SS_W.$$

The degrees of freedom are given in Table 6.4.

Table 6.4 Degree of freedom.

Degree of freedom	Total	Lot	Wafer	Measurement
	$df_{\text{tot}} = N - 1$	$df_L = L - 1$	$df_w = LW - 1$	$df_M = LW(M - 1)$

Table 6.5 ANOVA table for oxide thickness.

Source	df	SS	MS
Measurements	84	10 155.00	120.89
Wafers	63	36 014.50	571.66
Lots	20	61 442.29	3072.11
Total	167	107,611.79	

The mean squares for lots, wafers, and measurements are respectively,

$$MS_L = \frac{SS_L}{df_L}, \quad MS_W = \frac{SS_W}{df_W}, \quad MS_M = \frac{SS_M}{df_M}.$$

The variance components are given by

$$\sigma^2_{\text{meas}} = MS_M,$$
$$\sigma^2_{\text{wafers}} = \frac{MS_W - MS_M}{M},$$
$$\sigma^2_{\text{lots}} = \frac{MS_L - MS_M}{WM}.$$

The ANOVA table for the oxide thickness is given in Table 6.5.

REFERENCES

1. Harry, M. and Schroeder, R. (2000) *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations*. New York: Doubleday.
2. Montgomery, D.C. (2005) *Introduction to Statistical Quality Control*, 5th edition. Hoboken, NJ: John Wiley & Sons, Inc.
3. Pearson, K. (1894) Contributions to the mathematical theory of evolution I. On the dissection of asymmetrical frequency curves. *Philosophical Transactions of the Royal Society of London A*, 185, 71–110.
4. Mood, A.M., Graybill, F.A., and Boes, D.C. (1974) *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
5. Alwan, L.C. (2000) *Statistical Process Analysis*. Boston: Irwin/McGraw-Hill.

7

Fishbone Diagrams vs. Mind Maps

Timothy Yoap

Mind maps were developed by Tony Buzan in 1970. They originated from research into how the brain works and how to maximize its use. They are very similar to the cause-and-effect diagram used in measure phase of a Six Sigma project. The mind mapping technique has been introduced in many schools and educational institutions. In this chapter, similarities and differences, pros and cons as well as suggestions on how to incorporate the mind map technique into the Six Sigma toolbox are discussed.

7.1 INTRODUCTION

Cause-and-effect diagrams were developed by Kauro Ishikawa of Tokyo University in 1943, and thus are often called *Ishikawa diagrams*. They are also known as *fishbone diagrams* because of their appearance (in the plotted form). Cause-and-effect diagrams are used to systematically list the different causes that can be attributed to a problem (or an effect). A cause-and-effect diagram can aid in identifying the reasons why a process goes out of control.

Mind maps, also known as 'the Swiss Army knife of the brain', were developed by Tony Buzan in 1970.[1] They originated from research into how the brain works and how to maximize its use. Therefore, the original intent of the tool was to enhance learning, memory and creative thinking through simulation of how the brain works.

Although the two were developed in different times and places, and with different objectives in mind, they share some similarities and appear to be a good substitutes in many circumstances. On the other hand, because of those different objectives, there are differences between them that are too important to be ignored without degrading the power of the tools.

In this chapter we will start by going through a step-by-step procedure for creating a mind map.[2] This will be followed by a discussion of the similarities and differences

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

and the advantages and disadvantages of the mind map and its closest competitor in the Six Sigma toolbox in the context of how they can be used in the measure phase of the DMAIC Six Sigma Methodology. Finally, recommendations will be given on how mind maps can be integrated into the Six Sigma methodology, including the design for Six Sigma arena.

7.2 THE MIND MAP STEP BY STEP

The mind map can be created by the following steps. Throughout the whole map, the use of colours and images is highly recommended as these will stimulate the use of the right-hand side of the brain to enhance memory and creativity. A simple example of planning a public speech will be used to illustrate the creation of the mind map.

1. Create the central idea (in word or image) in the centre of the page. The central idea should illustrate clearly what the topic is all about (see Figure 7.1). Make sure the central word or image is big enough (approximately 2 square inches for a A4 paper size).
2. Add main branch to the central idea. The main branch is a curved line radiating out from and touching the central image, thicker towards the central image and thinner away from it. This actually simulates how ideas are radiated out from the brain.
3. Add a word on the main branch. The main branch should have a word or image "sitting" on the branch to represent one of the main ideas about the topic. The original recommendation was strictly one word per branch. The word should be in capital letters, and follow the curvature of the branch (see Figure 7.2). Today, many Six Sigma practitioners find that a little flexibility and slight bending of the rules actually make creating the mind map much easier without affecting its abilility to fulfill its purpose. For example, short phrases also work well. While a single word may help in memorizing the map, a short phrase actually has the advantage

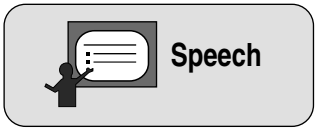


Figure 7.1 Mind map: central Idea.

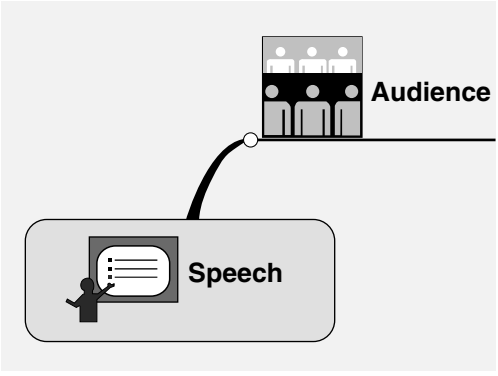


Figure 7.2 Mind map: main branch.

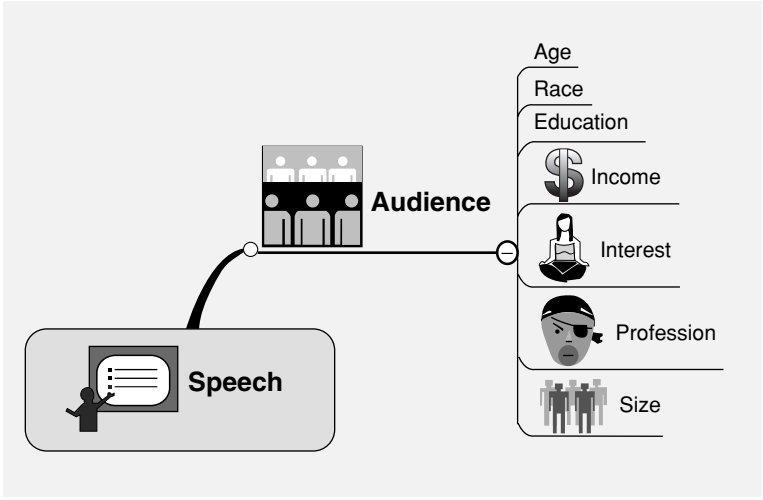


Figure 7.3 Mind map in progress.

of clearer communication. Furthermore, commercial software is available to make creating a mind map very simple, but depending on the software used, some of the original rules, which are supposedly ideal for enhancing memory and triggering creativity, may be violated.

4. Add words or images off the main branch. Adding the next level branches representing the ideas triggered by the main branch. The rules are essentially the same as for the main branch, except that the branch needs to be thinner, and the words and images smaller and less prominent than for the main branch.
5. Continue step 4 and keep expanding (see Figure 7.3).
6. Work through steps 2 to 5 again until the stock of ideas is exhausted (see Figure 7.4).

7.3 COMPARISON BETWEEN FISHBONE DIAGRAMS AND MIND MAPS

7.3.1 Similarities

As shown in the sample diagrams in Figure 7.5 and 7.6, both the fishbone diagram and mind map are able to capture and present a huge amount of ideas in one single page. They can both be used to facilitate brainstorming and help trigger new ideas generation.

7.3.2 Differences

The key difference between the fishbone diagram and the mind map is the appearance of the diagram. It is believed that the appearance of the diagram is able to direct and influence the thinking process.

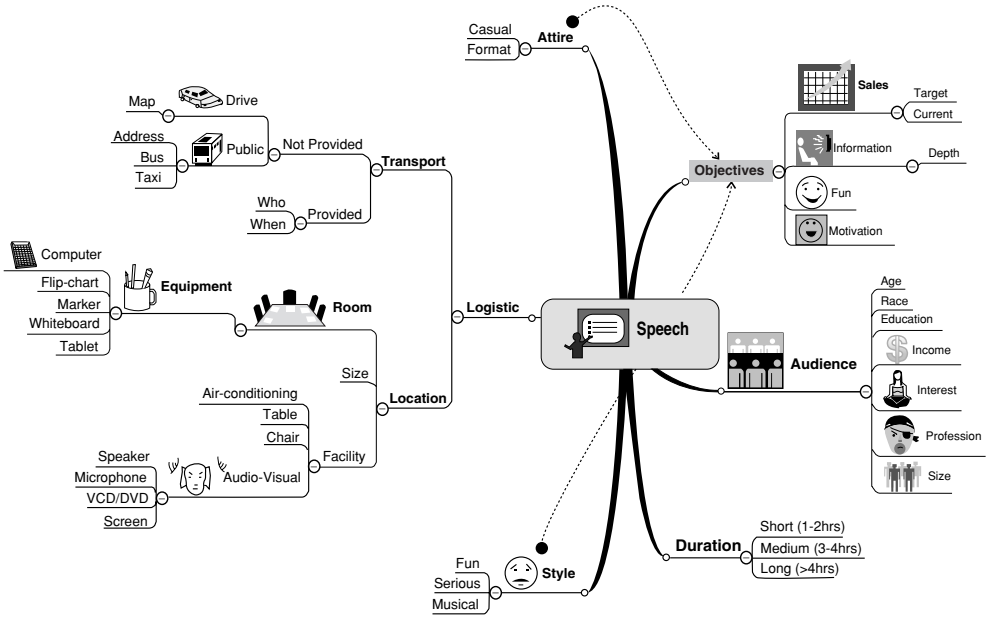


Figure 7.4 Completed mind map.

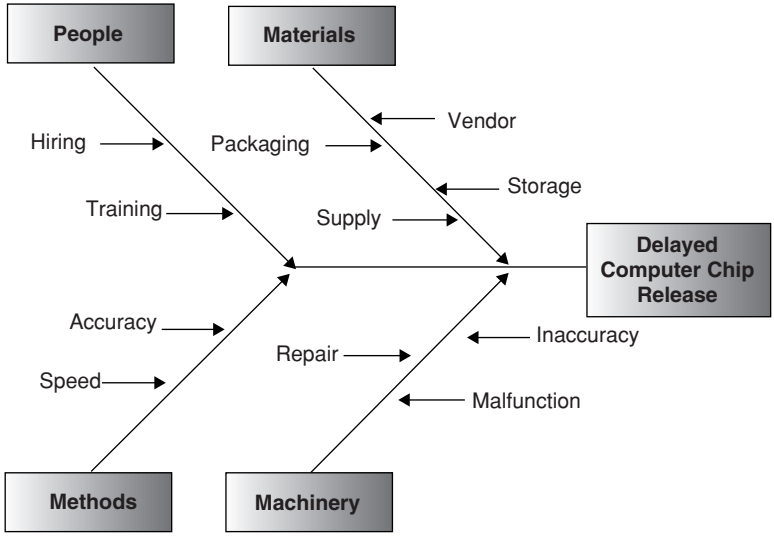


Figure 7.5 Sample fishbone diagram.

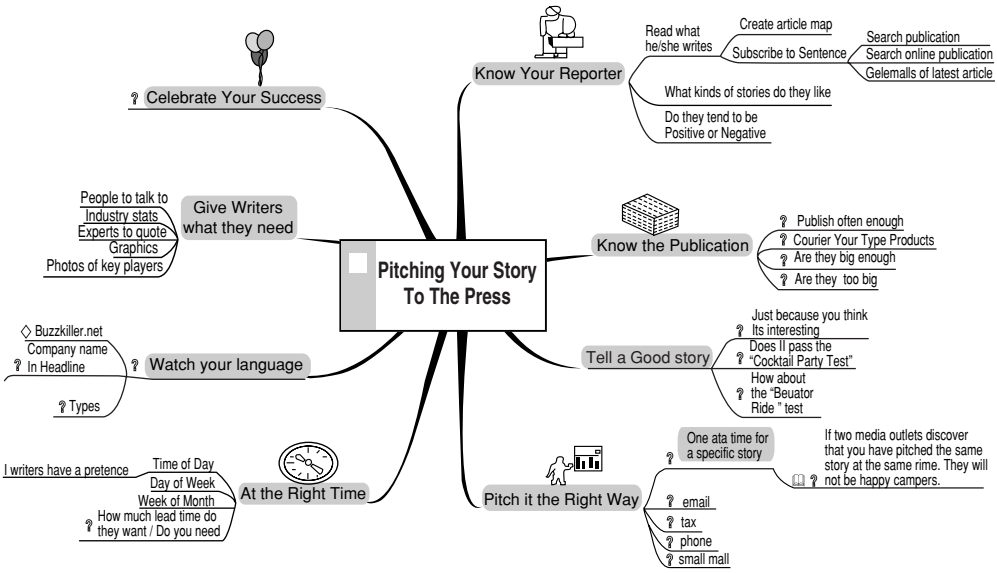

Pitching Your Story To The Press
Celebrate Your Success
Know Your Reporter
Read what he/she writes
Create article map
Subscribe to Sentence
Search publication
Search online publication
Gelemalls of latest article
What kinds of stories do they like
Do they tend to be Positive or Negative
Know the Publication
Publish often enough
Courier Your Type Products
Are they big enough
Are they too big
Tell a Good story
Just because you think
Its interesting
Does II pass the "Cocktail Party Test"
How about the "Beuator Ride " test
Pitch it the Right Way
One ata time for a specific story
If two media outlets discover that you have pitched the same story at the same rime. They will not be happy campers.
email
tax
phone
small mall
At the Right Time
I writers have a pretence
Time of Day
Day of Week
Week of Month
How much lead time do they want / Do you need
Watch your language
Buzzkiller.net
Company name
In Headline
Types
Give Writers what they need
People to talk to
Industry stats
Experts to quote
Graphics
Photos of key players

Figure 7.6 Sample mind map.

7.3.2.1 Direction of arrows

For the fishbone diagram, the arrows are pointed towards the 'effect' (as shown in Figure 7.7), which in our case is the problem we are trying to solve. Thus, a brainstorming session using this diagram will tend to be very focused. Every new idea (or 'cause') generated is likely to have direct relationship with the effect.

As for the mind map, there are no arrows pointing in any direction, but lines radiating out from the center of the page (as shown in Figure 7.8). This kind of diagram will inevitably direct the thinking into 'exploration' mode and, not uncommonly, lose sight of the 'shore' (the original intent).

7.3.2.2 Colour, line thickness and symbols (or images)

For the fishbone diagram, there is not much emphasis on the use of different colours and line thicknesses to differentiate the different levels or signal differences in

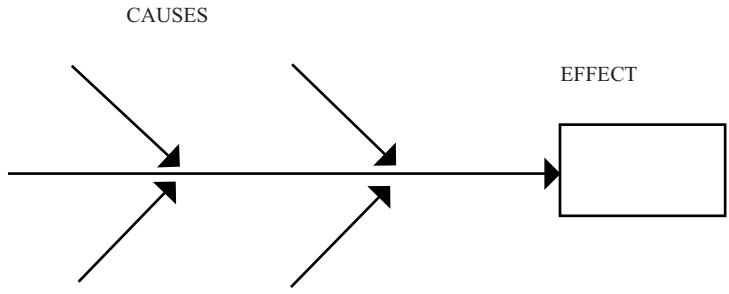


Figure 7.7 Direction of arrows for fishbone diagram.

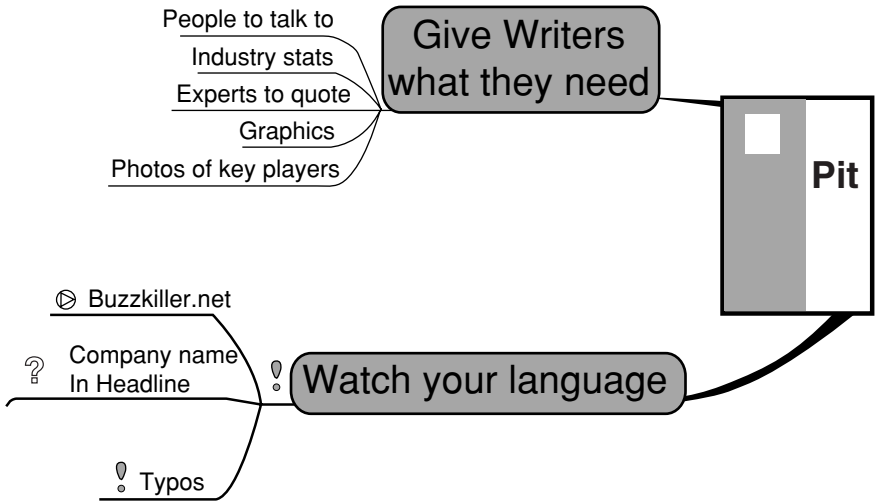


Figure 7.8 Direction of branches for mind mapping.

importance. This is understandable as the primary function of the tool is just to trace the cause-and-effect relationship and present it in diagram form. Symbols (C, N, and X) to indicate whether a key process input variable is a 'controllable', 'noise' or design factor may added to indicate the nature of the input variables. For the mind map, the use of colors, line thickness and images is almost mandatory. The reason is that the use of all these will stimulate the senses to improve the efficiency of the brain in memorizing them and trigger the creative process.

7.3.2.3 Guidelines for the main branches

The established guidelines for the fishbone diagram – man, machine, method, measurement, material and environment (M^5 E) for operational scenarios, and policy, people procedure, place, measurement and environment (P^4ME) for transactional scenarios – have been proven to be good guidelines that trigger the team to start thinking in a certain direction. As for the mind map, no such guideline exists, as, firstly, creativity is one of the key objectives of the tool. A guideline like this may stifle creativity as human beings generally like to take the easy way out and just limit their thinking to a small number of few categories. Secondly, the intended application of the mind map is too wide to give a sensible, general guideline for the main branches.

7.3.3 Other practical considerations

7.3.3.1 Ease of use and software support

Ease of use and software support are a key concern when a new tool is being introduced. For example, we could not implement Six Sigma fully if there were no statistical software and computer power available in the market.

The fishbone diagram has the advantage that it can be created using common software such as PowerPoint, Excel, Visio or any other graphic software. On the other hand, the mind map has the advantage of having powerful dedicated software that makes creating a mind map effortless, although without it too much effort is required to create a mind map the way it should look.

7.3.3.2 Acceptance level

All new tools need time before they are accepted. The fishbone diagram has the advantage of being the industrial standard as a brainstorming tool for problem-solving and for quality improvement. The mind map, on the other hand, is still not widely accepted in industry although it has been gaining popularity over the past few years. What will probably close the gap between them is that for a person who is familiar with the fishbone diagram, learning about the mind map does not take much effort.

7.4 CONCLUSION AND RECOMMENDATIONS

Looking at the tools from objective point of view in the context of Six Sigma, what we need in the measure phase is a tool that enables us to brainstorm the possible causes of a problem effectively and to summarize them in a one-page diagram. The fishbone diagram, which has served the Six Sigma community well for many years, remains a good enough tool. Nevertheless, we can also capitalize on the software support and the useful ideas of the mind map. Below are some recommendations:

1. Introduce the mind map as alternative way of presenting the fishbone diagram for those who have mind mapping software. This will improve the efficiency of creating fishbone diagrams and make them more presentable.
2. Include the mind map as a tool in define/measure phase when the problem is not very well defined.
3. Include the mind map as one of the tools for design for Six Sigma. Creating a new design requires mapping of the customer's needs. The mind map can act as a more effective input generator for quality function deployment than the fishbone diagram.
4. When using mind maps as the replacement for fishbone diagrams, use the established guidelines for the fishbone diagram – M^5E for operational and P^4ME for transactional – so as not to miss out any important area and make the brainstorming session more focused and efficient.

REFERENCES

1. Buzan, A. (1996) *The Mind Map Book: How to Use Radiant Thinking to Maximize Your Brain's Untapped Potential.* Toronto: Plume.
2. Mukerjea, D. (1998) *Superbrain.* Singapore: Oxford University Press.

8

Current and Future Reality Trees

Timothy Yoap

Traditional Six Sigma tools are very effective for addressing operational issues where problems are well-defined (e.g. reducing the failure rate of a certain failure mode), key process input and output variables are measurable, and solutions are arrived at by reducing variation through optimization of input parameters. In the transactional and business worlds, although there are attempts to force-fit or adapt the existing tools to solve problems, most of the time they are either not efficient or totally ineffective. The traditional Six Sigma toolset is often insufficient, especially in defining the problem and exploring probable solutions. In this chapter, we present the current reality tree and the future reality tree, two of the six thinking process tools in the theory of constraints developed by Goldratt. These perform similar functions to traditional Six Sigma tools. We compare the various tools and suggest how to incorporate current and future reality trees into the Six Sigma toolset.

8.1 INTRODUCTION

The current reality tree (CRT) and future reality tree (FRT) are two of the thinking process tools in the theory of constraints (TOC), a management philosophy developed by Eliyahu M. Goldratt.[1,2] The TOC is divided into three key interrelated areas: performance measurement, logistic control, and logical thinking processes. Performance measurement includes throughput, inventory, and operating expenses, and is concerned with making sure resources are focused on the bottleneck. Logistic control includes drum–buffer–rope scheduling and buffer management. Thinking process tools are important in understanding the current situation and eventually identifying the root problem (using the CRT), identifying and expanding win–win solutions (using the FRT and evaporating cloud), and developing implementation plans (using the prerequisite tree and transition tree).

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

The CRT is a logic-based tool for using cause-and-effect relationships to determine root problems that cause the undesirable effects in a given system. Its purpose is to depict the state of reality as it currently exists in the system. We often think that we know the exactly what the problem is, but what we see are only the effects of the problem and not the true problem itself. This is especially true in transactional projects where problems are not well defined. The traditional Six Sigma toolset is lacking in this area to help in defining a project.

The FRT is another logic-based tool, used after the CRT for constructing and testing potential solutions before implementation. The objectives are to develop, expand, and complete the solution, and identify and solve or prevent new problems created by implementing the solution.

In this chapter, we will start by going through the step-by-step procedure for creating a CRT and an FRT. This will be followed by a discussion of the similarities and differences and the advantages and disadvantages of the CRT, FRT and their competitors – such as failure mode and effect analysis (FMEA) and quality function deployment (QFD)[3] – in the Six Sigma toolset.

8.2 CURRENT REALITY TREE

In this section, we will show the steps involved in constructing an effective CRT. Note that this is not the only way this can be done. It is merely one way that has a simple structure and is easy for beginners to follow. With more experience, users can develop their own methods to simplify the process. A simple example, concerned with customer satisfaction with a telephone helpdesk for credit card services, is used to illustrate the steps.

The first step is to define the system boundaries and goals. This is to make sure the project is of the right size and to avoid suboptimization. Multiple goals are acceptable and encouraged to make sure the complete function of the system is captured. In the case of our example, the boundary is the telephone helpdesk for credit card services, and the goal is to improve customer satisfaction.

The next step is to identify any undesirable effects (UDEs) and preliminary causes. UDEs with respect to the goal of the system are identified by brainstorming the *causes*, *negatives*, and *whys* (in that order).

Negatives can be any indicators that tell you the performance of the system is not up to expectations. List these negatives (typically between five and ten will be sufficient) and name them N1, N2 . . . (see Table 8.1).

Table 8.1 List of negatives.

Negatives	
N1.	Operators cannot cope with the number of calls
N2.	User waste time going through all the options
N3.	Most of the time still need to talk to operator

Table 8.2 List of negatives and corresponding whys.

Negatives	Whys
N1. Operators cannot cope with the number of calls	W1. User feels frustrated W1a. User waits too long to be served
N2. User wastes time going through all the options	W2. User feels frustrated W2a. User wastes time on non-value-added waiting
N3. Most of the time still need to talk to operator	W3. Wasted time going through all the other options W3a. Need to wait a long time for the operator to attend to the call

The whys explain why the negatives are bad in relation to the goal. There can be more than one why for any negative. This is to justify the importance of the negatives. If you cannot say why a given 'negative' is bad in relation to the goal of the system, it cannot be that important. Typically, all whys are UDEs by definition. Write down the whys to the right of the negatives, naming them in an analogous manner to the negatives, so that whys W1, W1a, W1b, ... are associated with N1, and W2, W2a, ... with N2 (Table 8.2).

The causes are the reasons why the negatives have occurred, and there can be more than one for any negative. Write these down to the left of the negatives, numbering them in the same way as the whys (Table 8.3). Notice that the whole table is quite messy; some whys are the same as the negatives in another row, and some negatives are actually causes. This is to be expected, as there is no fixed pattern of the cause-and-effect relationship. This table is just the first brainstorming stage to gather enough

Table 8.3 List of negatives and corresponding whys and causes.

Causes	Negatives	Whys
C1. Pre-set option not comprehensive enough C1a. Operators are inexperienced C1b. Not enough operators	N1. Operators cannot cope with the number of calls	W1. User feels frustrated W1a. User waits too long to be served.
C2. Options are not arranged in order of popularity C2a. Classifications are not clear	N2. User wastes time going through all the options	W2. User feels frustrated W2a. User wastes time on non-value-added waiting
C3. Users do not want to go through the options C3a Users do not want to deal with a machine C3b. Users cannot get help from the options	N3. Most of the time still need to talk to operator	W3. Wasted time going through all the other options W3a. Need to wait a long time for the operator to attend to the call

Figure 8.1 Flowchart shapes used in drawing up a CRT.

information to kick-start the construction of the CRT. Reorganization of the ideas is required every now and then throughout the whole process.

The third step is to translate the items into CRT entities. These are flowchart shapes. The commonly used CRT entities are shown in Figure 8.1.

First, assign the UDEs. Recall that typically all whys are UDEs by definition (Figure 8.2). Then add the entities to the UDEs as shown in Figure 8.3.

Next, rearrange the boxes to remove redundancy. In Figure 8.3, W3 and N2 are the same and one of them can be removed. At this point, you may feel that 'User wasted time waiting to be served' could be one of the reasons why user feels frustrated, and an arrow can be added to show the cause-and-effect relationship. Also you may realize that 'Operators cannot cope with the number of calls' probably would not cause 'User feels frustrated' but affect the waiting time, so the arrow can be shifted to point to the right box. The result is shown in Figure 8.4.

Now select a branch and start expanding it. Add causes to each branch, refine the statement if needed, and test the cause–effect relationship using if–then statements. If the cause–effect relationship is insufficient, add the ellipse connector (if two or more causes need to happen together to bring about the effect) or additional condition (Figure 8.5),

Next re-examine the CRT, asking the following questions:

1. Are the UDEs appropriate and sufficient? That is, are they bad enough to be called UDEs, and will addressing them successfully enable the goal to be reached? Remove any inappropriate UDEs and add new ones if necessary.
2. Are the cause–effect relationships real and sufficient? That is, if the cause happened, then will the effect follow? Feel free to move the entities around and add ellipses if needed.
3. Might anything at a higher level make worse a cause–effect statement at a lower level? If so, add a negative reinforcing loop. In Figure 8.6, operators cannot cope with the workload which will likely cause higher turnover rate, so a negative reinforcing loop is added.

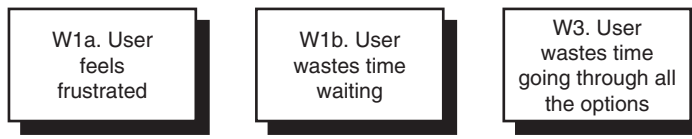


Figure 8.2 Assignment of undesirable effects.

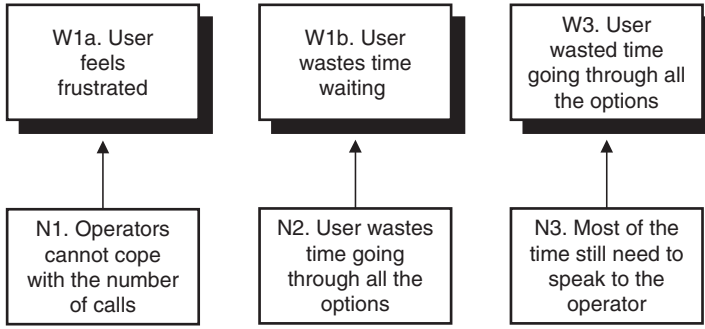

W1a. User feels frustrated
W1b. User wastes time waiting
W3. User wasted time going through all the options
N1. Operators cannot cope with the number of calls
N2. User wastes time going through all the options
N3. Most of the time still need to speak to the operator

Figure 8.3 Undesirable effects and CRT entities.

Now identify root causes for the UDEs and number the entities for easy reference (Figure 8.7). Typically, entities with no arrows going into them are root causes. Try to follow some standard rules for the numbering, especially if the CRT extends beyond one page. For example, the first page might start at 100, the second at 200, and so on.

At this point the CRT is complete. All the possible root causes have been identified. Note that you do not have to solve all the root causes to remove the UDEs. The next step is to construct an FRT to identify and evaluate the possible solutions.

8.3 FUTURE REALITY TREE (FRT)

The FRT is a tool for revealing how changing the current status can affect the future 'reality'. It is most commonly used to identify solutions (*injections* in TOC terminology) to problems identified during the construction of a CRT. It also allows evaluation of the injections before resources are committed to them. In this section, we show how to construct an FRT based on the example from the previous section.

The first step is to identify the desirable effects. This is typically done by stating the opposite of the UDEs from the CRT (Figure 8.8).

The second step is to list the possible injections for the root causes (Figure 8.9).

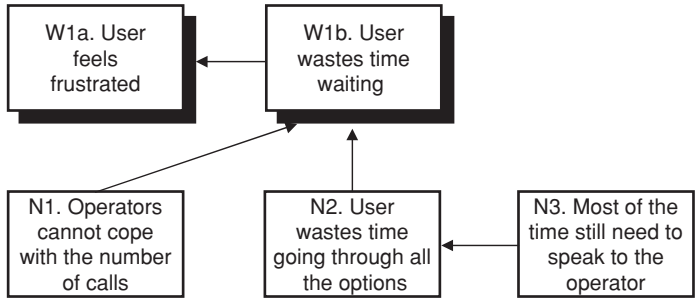

W1a. User feels frustrated
W1b. User wastes time waiting
N1. Operators cannot cope with the number of calls
N2. User wastes time going through all the options
N3. Most of the time still need to speak to the operator

Figure 8.4 Rearrangement of CRT to remove redundancy.

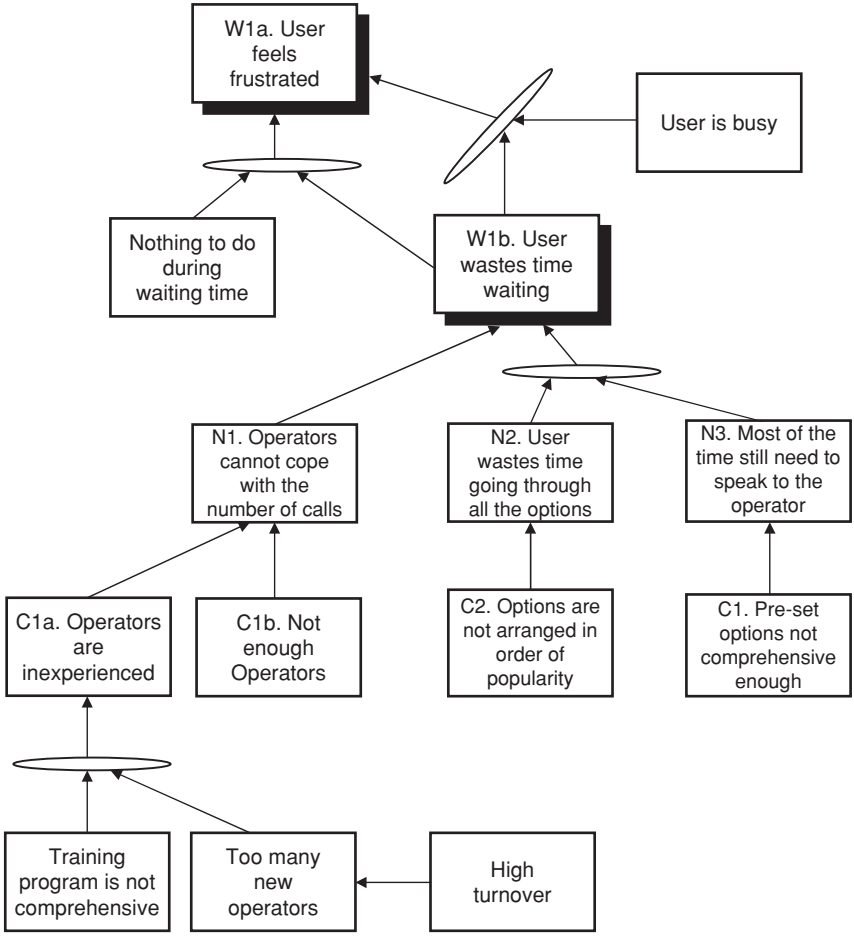

W1a. User feels frustrated
User is busy
Nothing to do during waiting time
W1b. User wastes time waiting
N1. Operators cannot cope with the number of calls
N2. User wastes time going through all the options
N3. Most of the time still need to speak to the operator
C1a. Operators are inexperienced
C1b. Not enough Operators
C2. Options are not arranged in order of popularity
C1. Pre-set options not comprehensive enough
Training program is not comprehensive
Too many new operators
High turnover

Figure 8.5 Expanding a branch.

The third step is to insert the injections into the CRT and change the effects where appropriate. Add new injections if needed until they are sufficient to reach the desirable effects (Figure 8.10).

The fourth and final step is to look for possible negative effects due to the introduction of the injections. If there are any, brainstorm new injections that can take care of them. This is to assess the feasibility of the solutions and also, to a certain extent, to error-proof them. If there is no good solution to the new negative effects, and the particular injection is not the only way to achieve the desirable effect, then it may be a good idea to drop the injection. As shown in Figure 8.11, adding new operators will increase operating costs; if there is no direct solution to this, it might be desirable to drop the injection, especially if the operators are felt to be well trained and efficient.

The FRT is now complete. The next step is to implement the solutions. This step can vary from an easy fix such as hiring two more operators, to forming a project team to revamp the entire training program.

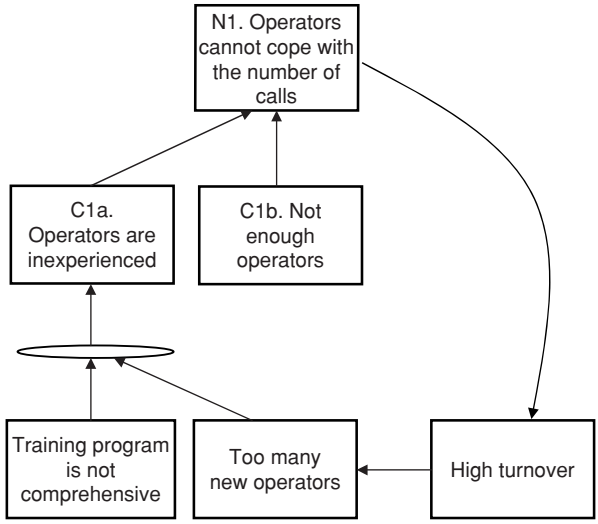

N1. Operators cannot cope with the number of calls
C1a. Operators are inexperienced
C1b. Not enough operators
Training program is not comprehensive
Too many new operators
High turnover

Figure 8.6 Addition of a negative reinforcing loop.

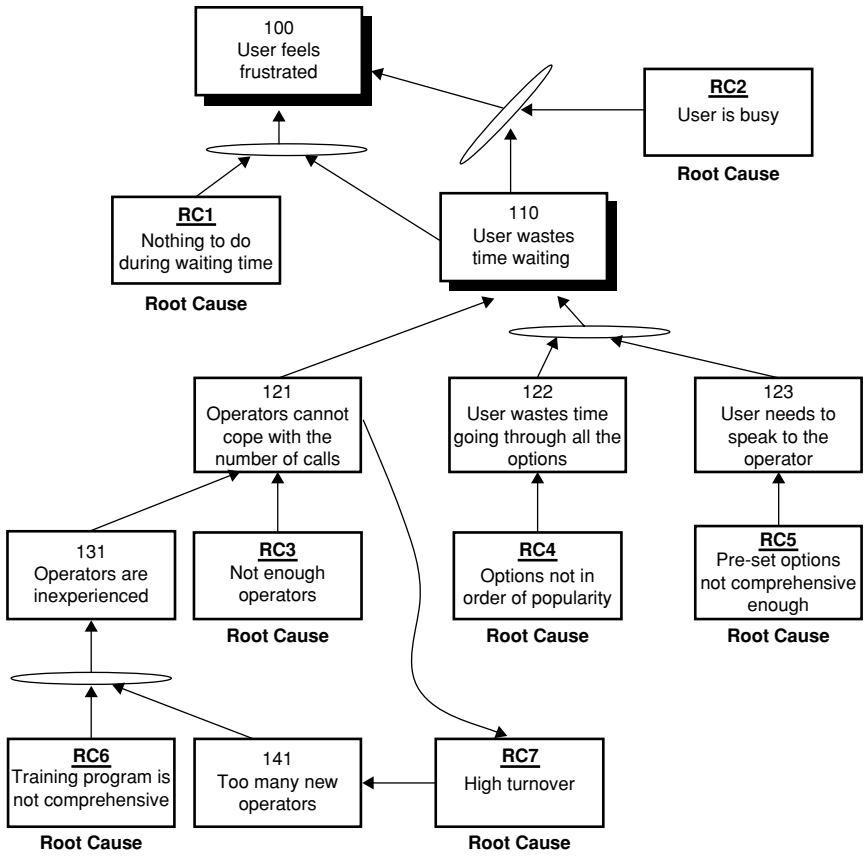


Figure 8.7 Addition of root causes.

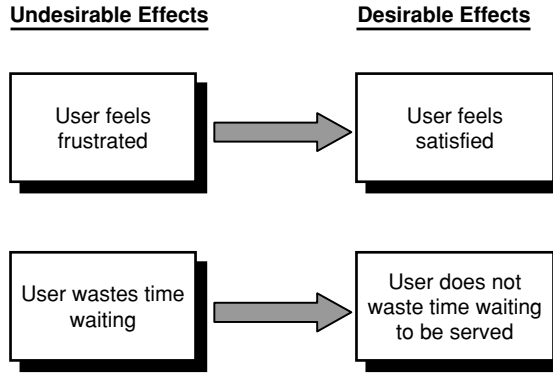


Figure 8.8 Desirable effects.

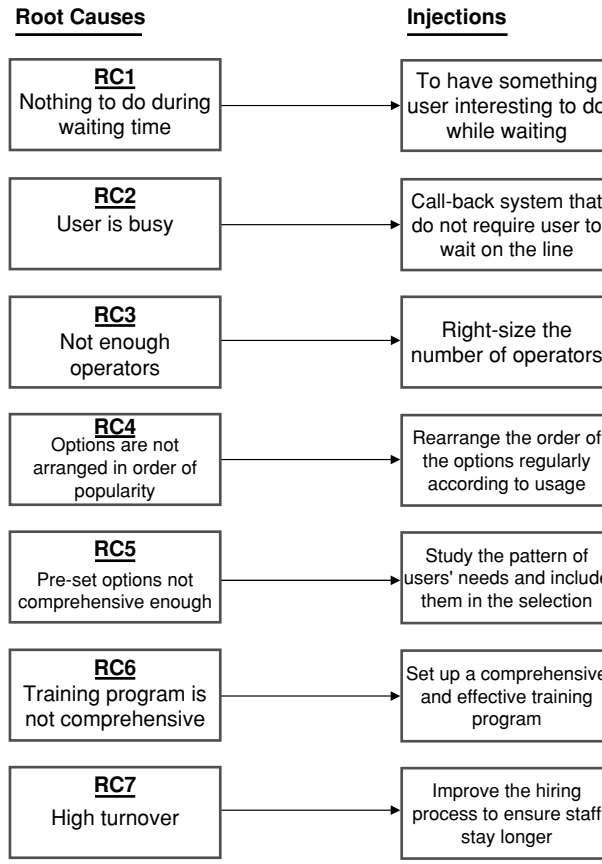

Root Causes
Injections
RC1
Nothing to do during waiting time
To have something user interesting to do while waiting
RC2
User is busy
Call-back system that do not require user to wait on the line
RC3
Not enough operators
Right-size the number of operators
RC4
Options are not arranged in order of popularity
Rearrange the order of the options regularly according to usage
RC5
Pre-set options not comprehensive enough
Study the pattern of users' needs and include them in the selection
RC6
Training program is not comprehensive
Set up a comprehensive and effective training program
RC7
High turnover
Improve the hiring process to ensure staff stay longer

Figure 8.9 Root causes and their injections.

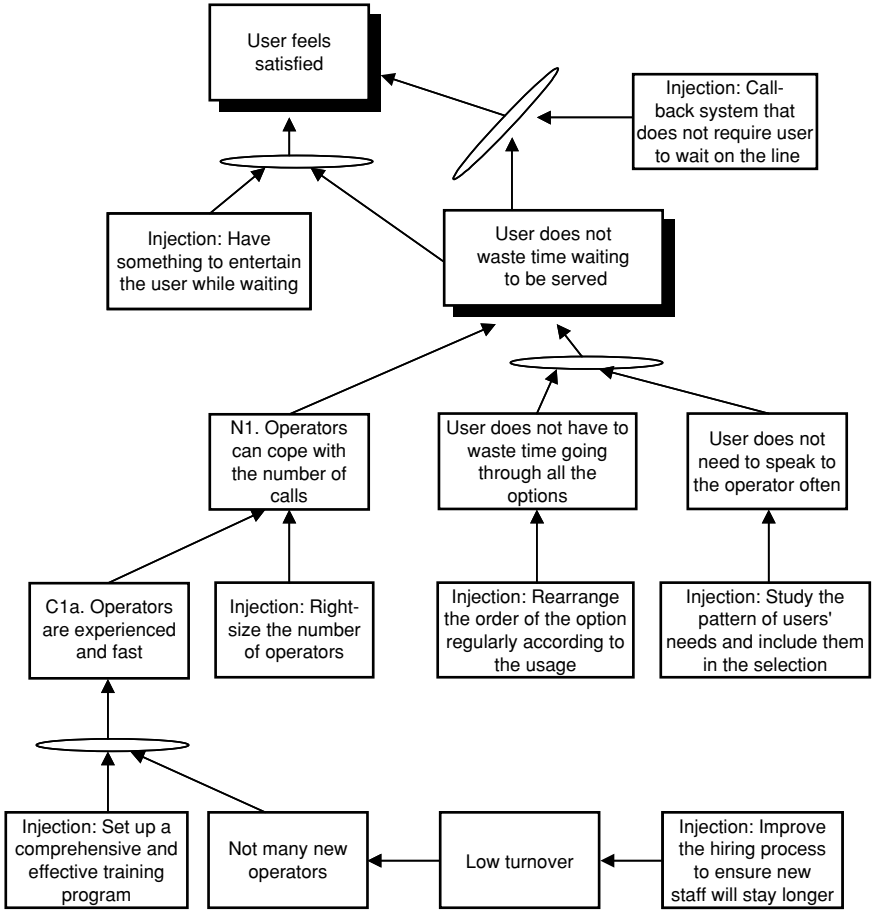

User feels satisfied
Injection: Call-back system that does not require user to wait on the line
Injection: Have something to entertain the user while waiting
User does not waste time waiting to be served
N1. Operators can cope with the number of calls
User does not have to waste time going through all the options
User does not need to speak to the operator often
C1a. Operators are experienced and fast
Injection: Right-size the number of operators
Injection: Rearrange the order of the option regularly according to the usage
Injection: Study the pattern of users' needs and include them in the selection
Injection: Set up a comprehensive and effective training program
Not many new operators
Low turnover
Injection: Improve the hiring process to ensure new staff will stay longer

Figure 8.10 CRT with injections added.

8.4 COMPARISON WITH CURRENT SIX SIGMA TOOLS

Are there any benefits in incorporating the CRT and FRT into the Six Sigma curriculum, and if so, how? As the scope of the CRT and FRT overlaps with some of the existing Six Sigma tools – root cause analysis (RCA), the cause-and-effect (C&E) matrix, FMEA, QFD, and design of experiments (DOE) – is it better to stick to one tool or to have alternative tools for people to choose? In my view it is better to stick to one and use it well. Using a common platform can only improve communication and effectiveness. With the proliferation of tools with overlapping uses, deciding when to use what is a serious problem for Black Belts.

Although some of these tools were created as a 'complete system' with their own purposes (e.g. FMEA and RCA), they have been adopted into Six Sigma as just a component in the DMAIC process. Thus, the following comparison is confined to the usage of tools within the Six Sigma environment.

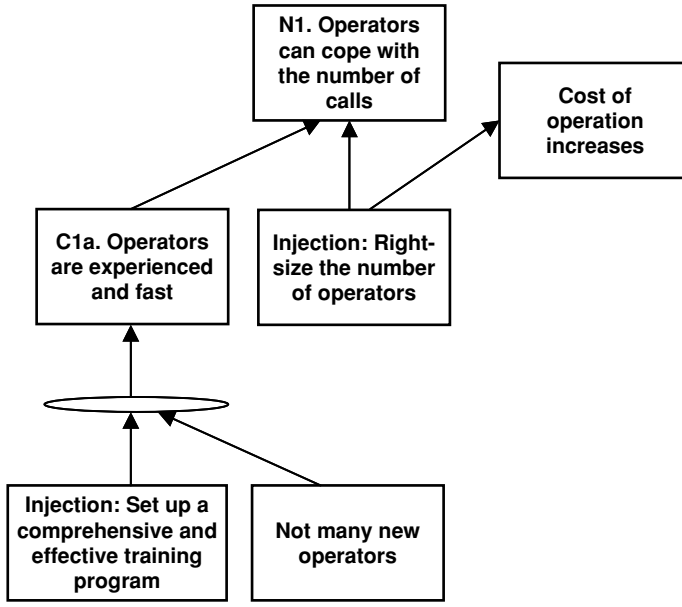

N1. Operators can cope with the number of calls
Cost of operation increases
C1a. Operators are experienced and fast
Injection: Right-size the number of operators
Injection: Set up a comprehensive and effective training program
Not many new operators

Figure 8.11 Identification of new undesirable effects resulting from injections.

8.4.1 Comparison of CRT, RCA, FMEA and C&E matrix

The four current Six Sigma tools are compared in Table 8.4. The objective is to organize and understand the similarities and differences between them.

8.4.2 CRT and RCA

RCA is the most direct 'competitor' of the CRT. Both employ the cause–effect thinking process and produce similar-looking diagrams .The purpose of CRT is to search for the 'root causes' of a list of 'Negative effects'. The objective of RCA is to trace the 'root causes' of a given problem. The differences between them are generally subtle, but may be significant at times. In the Six Sigma strategy, these are supposed to be in different phases, the define phase for the CRT and measure phase for RCA.

8.4.3 FRT vs. QFD, FMEA and DOE

Objectively, we need a tool in the improve phase in order, firstly, to identify the solutions to the problem and, secondly, design a solution if necessary. In the traditional transactional Six Sigma curriculum, the job is taken care of partly by FMEA (identify needed solution) and partly by DOE (optimize a process). From experience, FMEA is not very popular unless it is enforced by management as a system. As for DOE, usage is even lower. Looking at this scenario, the FRT can be a good alternative for identifying the needed solutions. As for solution development, QFD and some relevant

Table 8.4 Comparison of current Six Sigma tools.

	CRT	RCA	FMEA	C&E Matix
Input	1) Negative effects of the problem 2) Customer complains	Problem statement	1) Problem statement 2) Fishbone diagram 3) C&E matrix	1) Problem statement 2) Fishbone diagram
Output	Suspected root causes of the negative effects	Root cause of a problem	Few most important KPIVs and some 2nd, 3rd and lower level causes	Few most important KPIVs
Prioritization	No	No	Yes	Yes
Summary	Suspected root causes of the negative effects	Root cause of a problem	1) List of most important causes (KPIVs) 2) Why they are important 3) What are the actions 4) When it is complete 5) Assessment after implementation	1) List of most important causes (KPIVs) 2) Why they are important (correlation to key requirement)
Require good knowledge to start?	Yes	Yes	No, but the more the better	No, but the more the better

(*Continued*)

Table 8.4 Comparison of current Six Sigma tools. (*Continued*)

	CRT	RCA	FMEA	C&E Matix
Facilitates grouping and regrouping	Yes	Yes	Yes, but not easy	No
Encourage 'data-driven' decision	Not really	Not really	Yes, long list of possible KPIVs to be verified with data	Yes, long list of possible KPIVs to be verified with data
Encourage asking 'why'	Yes	Yes	Not really, as it is harder to drill down to the next level with the FMEA Excel sheet	No
Acceptance level	1. New to Six Sigma, although not very difficult to get it accepted. 2. Some coverage of TOC in lean enterprise	Medium high	Medium high	Medium high
Software support	Visio, and other commercial software	Visio, and other commercial software	Excel	Excel

operations research techniques may be the answer. QFD allows the user to make sure that the solution developed fulfills all the requirements. Operations research involves the use of mathematical techniques to model and analyze the problem, so that the best (optimum) course of action, under the restriction of limited resources, may be determined.

8.5 CONCLUSION AND RECOMMENDATIONS

Current and future reality trees are powerful tools used as part of the under the total TOC package. They incorporate some useful ideas which we can borrow and work into the Six Sigma system. Some recommendations follow.

1. CRT is a good define phase tool for breaking problems down into a few well-defined root causes which we can use the normal Six Sigma tools and methodology to solve effectively. It is particularly effective for transactional projects where problems are not well defined.
2. The CRT is a good project selection tool and can even be included in Champion training. It is especially effective for transactional projects as it helps to drill down the negative effects experienced by Champions to one or more well-defined root causes. Each of the root causes can be a project.
3. Include the FRT in the improve phase as it can help in solution generation and in qualitatively checking the effectiveness of the solution. This will be of great help for projects that are impossible or inappropriate for such techniques as DOE.

REFERENCES

1. Goldratt, E.M. and Cox, J. (2004) *The Goal*. Great Barrington, MA: North River Press.
2. Goldratt, E.M. (1999) *Theory of Constraints*. Great Barrington, MA: North River Press.
3. Cohen, L. (1995) *Quality Function Deployment*. Reading, MA: Addison-Wesley.
4. Dettmer, H.W. (1997) *Goldratt's Theory of Constraints: A Systems Approach to Continuous Improvement*. Milwaukee, WI: ASQ Quality Press.
5. Dettmer, H.W. (2003) *Strategic Navigation: A Systems Approach to Business Strategy*. Milwaukee, WI: ASQ Quality Press.

9

Computing Process Capability Indices for Nonnormal Data: A Review and Comparative Study

L. C. Tang, S. E. Than and B. W. Ang

When the distribution of a process characteristic is nonnormal, the indices C_p and C_{pk}, calculated using conventional methods, often lead to erroneous interpretation of the process's capability. Though various methods have been proposed for computing surrogate process capability indices (PCIs) under nonnormality, there is a lack of literature offering a comprehensive evaluation and comparison of these methods. In particular, under mild and severe departures from normality, do these surrogate PCIs adequately capture process capability, and what is the best method for reflecting the true capability under each of these circumstances? In this chapter we review seven methods that are chosen for performance comparison in their ability to handle nonnormality in PCIs. For illustration purposes the comparison is carried out by simulating Weibull and lognormal data, and the results are presented using box plots. Simulation results show that the performance of a method is dependent on its capability to capture the tail behavior of the underlying distributions. Finally, we give a practitioner's guide that suggests applicable methods for each defined range of skewness and kurtosis under mild and severe departures from normality.

9.1 INTRODUCTION

Process capability indices (PCIs) are widely used to determine whether a process is capable of producing items within a specific tolerance. The most common indices, C_p

This chapter is based on the article by L. C. Tang and S. E. Than, 'Computing process capability indices for non-normal data: a review and comparative study', *Quality and Reliability Engineering International*, 15(5), 1999, pp. 339–353, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

and C_{pk}, are defined as follows:

$$C_p = \frac{\text{specification width}}{\text{process width}} = \frac{USL - LSL}{6\sigma}, \tag{9.1}$$

$$C_{pk} = \min\left\{C_{pu}, C_{pl}\right\} = \min\left\{\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right\}, \tag{9.2}$$

where USL and LSL denote the upper and lower specification limits, respectively.[1] Since the process mean μ and the process variance σ^2 are unknown, they are usually estimated from the sample statistics $\bar{x}$ and s^2.

English and Taylor[2] examined the effect of the nonnormality assumption on PCIs and concluded that C_{pk} is more sensitive to departures from normality than C_p. Kotz and Johnson[3] provided a survey of work on the properties of PCIs and their estimators when the distribution is nonnormal. Among them are Clements' method,[4] the Johnson–Kotz–Pearn method,[5] and 'distribution-free' PCIs.[6] A new index C_s, proposed by Wright,[7] incorporates an additional skewness correction factor in the denominator of the C_{pmk} index developed by Johnson *et al.*[5] Choi and Bai[8] proposed a heuristic weighted variance method to adjust the value of PCIs according to the degree of skewness by considering the standard deviations above and below the process mean separately.

Several authors have proposed a new generation of PCIs based on assumptions about the underlying population. Pearn and Kotz[9] based their new PCIs on a Pearsonian distribution. Johnson *et al.*[5] suggested replacing 6σ in the denominator of equation (9.1) by 6θ, where θ is chosen so that the "capability" is not greatly affected by the shape of the distribution. Castagliola[10] introduced a nonnormal PCI calculation method by estimating the proportion of nonconforming items using Burr's distribution. Vännman[11] proposed a new family of indices $C_p(u, v)$, parameterized by (u, v), which includes many other indices as special cases. Deleryd[12] investigated suitable values for u and v when the process distribution is skewed. $C_p(1,1)$, which is equivalent to C_{pmk}, is recommended as being most suited to handling nonnormality in PCIs. While it is well known that C_p and C_{pk} are not indicative of the process capability for nonnormal process characteristics and that various methods are available to compute some surrogate PCIs, there is a lack of a comprehensive performance comparison of these methods. In particular, practitioners are most interested in knowing which methods will give a PCI that can be compared, on the same scale, with C_{pk}, and how well the various methods perform under slight, moderate and severe departures from normality. In this chapter we examine the performance of seven different methods in reflecting the true capability of the process by comparing the C_{pu} value generated via simulations with a target C_{pu} value.

9.2 SURROGATE PCIs FOR NONNORMAL DATA

In this section a summary of the seven different methods included for comparison in this chapter is presented.

9.2.1 Probability plot

A widely accepted approach to PCI computation is to use a normal probability plot[13] so that the normality assumption can be verified simultaneously. Analogous to the normal probability plot, where the natural process width is between the 0.135 and 99.865 percentiles, surrogate PCI values may be obtained via suitable probability plots:

$$C_p = \frac{USL - LSL}{\text{upper } 0.135\% \text{ point} - \text{lower } 0.135\% \text{ point}}$$
$$= \frac{USL - LSL}{U_p - L_p},$$

where U_p and L_p are respectively the 99.865 and 0.135 percentiles of the observations. These percentile points can easily be obtained from the simple computer code that performs probability ploting. Since the median is the preferred central value for a skewed distribution, the equivalent C_{pu} and C_{pl} are defined as

$$C_{pu} = \frac{USL - \text{ median}}{x_{0.998\,65} - \text{ median}},$$
$$C_{pl} = \frac{\text{median } - LSL}{\text{median} - x_{0.998\,65}}.$$

C_{pk} is then taken as the minimum of (C_{pu},C_{pl}).

9.2.2 Distribution-free tolerance intervals

Chan *et al.*[6] adopted a distribution-free tolerance interval approach to compute C_p and C_{pk} for a nonnormal process using

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{USL - LSL}{\frac{3}{2}(4\sigma)}$$
$$= \frac{USL - LSL}{3(2\sigma)}.$$

This results in

$$C_p = \frac{USL - LSL}{w} = \frac{USL - LSL}{\frac{3}{2}w_2}$$
$$= \frac{USL - LSL}{3w_3},$$
$$C_{pk} = \frac{\min[(USL - \mu), (\mu - LSL)]}{w/2},$$

where w is the width of the tolerance interval with 99.73% coverage 95% of the time, w_2 is the width of the tolerance interval with 95.46% coverage 95% of the time, and w_3 is the width of the tolerance interval with 68.26% coverage 95% of the time. The order statistics estimates of w, w_2 and w_3, based on the normality assumption, are given by Chan *et al*.[6] This is a more conservative method, since the natural process width is greater as it is estimated taking the sampling variation into account. This is the only method considered here that will give a different result even if the underlying distribution is normal. It is included here to investigate whether its conservative nature is preserved under nonnormality.

9.2.3 Weighted variance method

Choi and Bai[8] proposed a heuristic weighted variance method to adjust the PCI values according to the degree of skewness of the underlying population. Let P_x be the probability that the process variable X is less than or equal to its mean μ,

$$P_x = \frac{1}{n}\sum_{i=1}^{n} I(\bar{X} - X_i),$$

where $I(x) = 1$ if $x > 0$ and $I(x) = 0$ if $x < 0$. The PCI based on the weighted variance method is defined as

$$C_p = \frac{USL - LSL}{6\sigma W_x}$$

where $W_x = \sqrt{1 + |1 - 2P_x|}$. Also

$$C_{pu} = \frac{USL - \mu}{2\sqrt{2P_x\sigma}},$$

$$C_{pl} = \frac{\mu - LSL}{3\sqrt{2(1 - P_x)\sigma}}.$$

9.2.4 Clements' method

Clements[4] replaced 6σ in equation (9.1) by $U_p - L_p$:

$$C_p = \frac{USL - LSL}{U_p - L_p},$$

where U_p is the 99.865 percentile and L_p is the 0.135 percentile.[14] For C_{pk}, the process mean μ is estimated by the median M, and the two 3σs are estimated by $U_p - M$ and $M - L_p$ respectively, giving

$$C_{pk} = \min\left[\frac{USL - M}{U_p - M}, \frac{M - LSL}{M - L_p}\right].$$

Clements' approach uses the classical estimators of skewness and kurtosis that are based on third and fourth moments respectively, which may be somewhat unreliable for very small sample sizes.

9.2.5 Box–Cox power transformation

Box and Cox[15] proposed a useful family of power transformations on the necessarily positive response variable X given by

$$X^{(\lambda)} = \begin{cases} \dfrac{X^{\lambda} - 1}{\lambda}, & \text{for } \lambda \neq 0, \\ 1\text{n } X, & \text{for } \lambda = 0. \end{cases} \tag{9.3}$$

This continuous family depends on a single parameter λ which can be estimated by the method of maximum likelihood as follows.

First, a value of λ from a selected range is chosen. For the chosen λ we evaluate

$$\begin{aligned} L_{\max} &= -\tfrac{1}{2} \ln \hat{\sigma}^2 + \ln J(\lambda, X) \\ &= -\tfrac{1}{2} \ln \hat{\sigma}^2 + (\lambda - 1) \sum_{i=1}^{n} \ln X_i, \end{aligned}$$

where

$$J(\lambda, X) = \prod_{i=1}^{n} \frac{\partial W_i}{\partial X_i} = \prod_{i=1}^{n} X_i^{\lambda - 1}, \qquad \text{for all } \lambda,$$

so that $\ln J(\lambda, X) = (\lambda - 1) \sum_{i=1}^{n} \ln X_i$. The estimate of $\hat{\sigma}^2$ for fixed λ is $\hat{\sigma}^2 = S(\lambda)/n$, where S($\lambda$) is the residual sum of squares in the analysis of variance of $X(\lambda)$. After calculating $L_{\max}(\lambda)$ for several values of λ within the range, $L_{\max}(\lambda)$ can be plotted against λ. The maximum likelihood estimator of λ is obtained from the value of λ that maximizes $L_{\max}(\lambda)$. With the optimal λ^* value, each of the X data specification limits is transformed into a normal variate using equation (9.3). The corresponding PCIs are calculated from the mean and standard deviation of the transformed data using equations (9.1) and (9.2).

9.2.6 Johnson transformation

Johnson[16] developed a system of distributions based on the method of moments, similar to the Pearson system. The general form of the transformation is given by

$$z = \gamma + \eta \tau(x; \varepsilon, \lambda), \qquad \eta > 0, -\infty < \gamma < \infty, \lambda > 0, -\infty < \varepsilon < \infty, \tag{9.4}$$

where z is a standard normal variate and x is the variable to be fitted by a Johnson distribution. The four parameters, γ, η, ε, and λ are to be estimated, and τ is an arbitrary function which may take one of the following three forms.

9.2.6.1 The lognormal system (S_L)

$$\tau_1(x; \varepsilon, \lambda) = \log\left(\frac{x - \varepsilon}{\lambda}\right), \qquad x \geq \varepsilon \tag{9.5}$$

This is the Johnson S_L distribution that covers the lognormal family. The required estimates for the parameters are

$$\hat{\eta} = 1.645\left[\log\left(\frac{x_{0.95} - x_{0.5}}{x_{0.5} - x_{0.05}}\right)\right]^{-1}, \tag{9.6}$$

$$\hat{\gamma}^* = \hat{\eta}\log\left(\frac{1 - \exp\left(-1.645/\hat{\eta}\right)}{x_{0.5} - x_{0.05}}\right), \tag{9.7}$$

$$\hat{\varepsilon} = x_{0.5} = \exp\left(-\hat{\gamma}^*/\hat{\eta}\right), \tag{9.8}$$

where the 100αth data percentile is obtained as the $\alpha(n + 1)$th-ranked value from n observations. If necessary, linear interpolation between consecutive values may be used to determine the required percentile.

9.2.6.2 *The unbounded system (S_U)*

$$\tau_2(x; \varepsilon, \lambda) = \sinh^{-1}\left(\frac{x - \varepsilon}{\lambda}\right), \qquad -\infty < x < \infty. \tag{9.9}$$

Curves in the S_U family are unbounded. This family covers the t and normal distributions, among others. For the fitting of this distribution, Hahn and Shapiro[17] gave tables for the determination of $\hat{\gamma}$ and $\hat{\eta}$ based on given values of kurtosis and skewness.

9.2.6.3 *The bounded system (S_B)*

$$\tau_3(x; \varepsilon, \lambda) = \log\left(\frac{x - \varepsilon}{\lambda + \varepsilon - x}\right), \qquad \varepsilon \le x \le \varepsilon + \lambda. \tag{9.10}$$

The S_B family covers bounded distributions, which include the gamma and beta distributions. Since the distribution can be bounded at either the lower end (ε), the upper end ($\varepsilon + \lambda$), or both, this leads to the following situations.

- Case I. *Range of variation known.* For the case where the values of both endpoints are known, the parameters are obtained as

$$\hat{\eta} = \frac{z_{1-\alpha'} - z_\alpha}{\log\left(\frac{(x_{1-\alpha'} - \varepsilon)(\varepsilon + \lambda - x_\alpha)}{(x_\alpha - \varepsilon)(\varepsilon + \lambda - x_{1-\alpha'})}\right)}, \tag{9.11}$$

$$\hat{\gamma} = z_{1-\alpha'} - \hat{\eta}\log\left(\frac{x_{1-\alpha'} - \varepsilon}{\varepsilon + \lambda - x_{1-\alpha'}}\right). \tag{9.12}$$

- Case II. *One endpoint known.* In this case an additional equation obtained by matching the median of the data is needed to supplement equations (9.11) and (9.12). This equation is given by

$$\hat{\lambda} = (x_{0.5} - \varepsilon)\left\{(x_{0.5} - \varepsilon)(x_\alpha - \varepsilon) + (x_{0.5} - \varepsilon)(x_{1-\alpha} - \varepsilon) - 2(x_\alpha - \varepsilon)(x_{1-\alpha} - \varepsilon)\right\} \times \left\{(x_{0.5} - \varepsilon)^2 - (x_\alpha - \varepsilon)(x_{1-\alpha} - \varepsilon)\right\}^{-1}.$$

- Case III. *Neither endpoint known.* For the case where neither endpoint is known, four data percentiles have to be matched with the corresponding percentiles of the standard normal distribution. The resulting equations for $i = 1, 2, 3, 4$,

$$z_i = \hat{\gamma} + \hat{\eta} \log\left(\frac{x_i - \hat{\varepsilon}}{\hat{\varepsilon} + \hat{\lambda} - x_i}\right),$$

  are nonlinear and must be solved by numerical methods.

The algorithm developed by Hill *et al.*[18] is used to match the first four moments of X to the above distribution families. PCIs are calculated using equations (9.1) and (9.2).

9.2.7 Wright's process capability index C_S

Wright[7] proposed a CPI, C_s, that takes into account the skewness by incorporating an additional skewness correction factor in the denominator of C_{pmk},[5] and is defined as

$$C_s = \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma^2 + (\mu - T)^2 + |\mu_3/\sigma|}} = \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma^2 + |\mu_3/\sigma|}},$$

where $T = \mu$ and μ_3 is the third central moment.

Some of the methods described above have been widely applied in industry, such as probability plotting and Clements' method; however, method such as the Box–Cox transformation is relatively unknown to practitioners. It should be noted that when the underlying distribution is normal, theoretically, all the above methods, with the exception of the distribution-free method, should give the same result as the conventional C_p and C_{pk} given in equations (9.1) and (9.2). Nevertheless, as they use different statistics and/or different ways of estimating the associated statistics, the resultant estimates for C_p and C_{pk} will exhibit some variability. In particular, the variability of C_p could be reduced with increasing sample size, as its sampling distribution is χ^2_{n-1}under the normality assumption. However, the variability of C_{pk} can be quite significant for all reasonable sample sizes, as it also depends on the variability in the process shift.[19]

9.3 SIMULATION STUDY

9.3.1 Yardstick for comparison

The very fact that different comparison yardsticks may well lead to different conclusions emphasizes the need to identify a suitable yardstick for comparing the performance of the above seven methods. In the literature, different researchers have utilized widely different yardsticks in tackling the nonnormality problem of PCIs.

English and Taylor[2] used fixed values of C_p and C_{pk} (equal to 1.0) for all their simulation runs in investigating the robustness of PCIs to nonnormality. The basis for their comparison was the proportion of $\hat{C}_p$ and $\hat{C}_{pk}$ (estimated from simulation) greater than 1.0 for the normal distribution case. Deleryd[12] focused on the proportion of

nonconforming items to derive the corresponding value of C_p. The bias and dispersion of the estimated $C_p(u, v)$ values are compared with the target C_p values. Rivera *et al.*[20] varied the upper specification limits of the underlying distributions to derive the actual number of nonconforming items and the equivalent C_{pk} values. Estimated C_{pk} values calculated from the transformed simulation data are compared with these target C_{pk} values.

In practice, PCIs are commonly used for tracking performance and comparison between different processes. Although such uses without examining the underlying distribution should not be encouraged, a "good" surrogate PCI for nonnormal data should be compatible with the PCI computed under normality when the corresponding fractions nonconforming are about the same. This motivates a scheme similar to that of Rivera *et al.*[20] where the fraction nonconforming is fixed a priori by using suitable specification limits, and PCIs computed using various methods are then compared with a target value. This leads to consideration of unilateral tolerance, where C_{pu}, a single tolerance limit capability index, is used as the comparison yardstick in our simulation study. For a targeted C_{pu} value the fraction of nonconforming units under the normality assumption can be determined using

$$\text{fraction nonconforming} = \Phi\left(-3C_{pu}\right). \tag{9.13}$$

In our simulation study, targeted values of $C_{pu} = 1$, 1.5 and 1.667 are used and the corresponding *USL* values for lognormal and Weibull distributions with the same fraction nonconforming are obtained. These *USL* values are then used to estimate the C_{pu} index pertaining to the different methods from the simulated data. These estimated C_{pu}s are then compared with the targeted C_{pu} values.

A superior method is one with its sample mean of the estimated C_{pu} having the smallest deviation from the target value (accuracy) and with the smallest variability, measured by the spread or standard deviation of the estimated C_{pu} values (precision). A graphical representation that conveniently depicts these two characteristics is the simple box-and-whisker plot.

9.3.2 Underlying distributions

Lognormal and Weibull distributions are used to investigate the effect of nonnormal data on the PCIs, as they are known to have parameter values that can represent slight, moderate, and severe departures from normality. These distributions are also known to have significantly different tail behaviors, which greatly influences the process capability. The sets of parameter values used in the simulation study are given in Table 9.1. To save space, detailed results are presented here only for those sets in bold type. The skewness and kurtosis for each of these distributions are given in Table 9.2.

The cumulative distribution functions (CDFs) of the lognormal and Weibull distributions are respectively given by

$$F\left(x; \mu, \sigma\right) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \qquad x \geq 0, \tag{9.14}$$

$$F\left(x; \eta, \sigma\right) = 1 - \exp\left[-\left(\frac{x}{\sigma}\right)^{n}\right], \qquad x \geq 0. \tag{9.15}$$

Table 9.1 Parameter combinations used in performance comparison simulation.

Lognormal		Weibull	
μ	σ^2	Shape	Scale
0.0	0.1	**1.0**	**1.0**
0.0	**0.3**	**2.0**	**1.0**
0.0	**0.5**	4.0	1.0
1.0	0.5	0.5	1.0

Figures 9.1 and 9.2 show respectively the shapes of the lognormal and Weibull distributions used in the simulation.

9.3.3 Simulation runs

A series of simulations was conducted with sample sizes $n = 50, 75, 100, 125$ and 150 and with target C_{pu} values of 0.1, 1.33, 1.5, 1.667 and 1.8 using lognormal and Weibull distributions. We also include results calculated using standard PCI expressions under the 'assumed normal' method. Here we present only the results for simulations with $n = 100$ and target C_{pu} values of 1.0, 1.5 and 1.667.

In each run the necessary statistics required by the different methods, such as $\bar{x}$, s, upper and lower 0.135 percentile and median, were obtained from the random variables generated from the respective distributions. For methods involving transformation, these statistics are essentially obtained from the transformed values. Estimates for C_{pu} were determined, using the above seven methods, from these estimated parameters for each of the targeted *USL* values.

Table 9.2 Skewness and kurtosis values of distributions investigated.

Lognormal distribution			
Mean	Variance	Skewness $\sqrt{\beta_1}$	Kurtosis (β_2)
0.0	0.5	2.9388	21.5073
0.0	0.3	1.9814	10.7057
0.0	0.1	1.0070	4.8558
1.0	0.5	2.9388	21.5073
Weibull distribution			
Shape	Scale	Skewness $\sqrt{\beta_1}$	Kurtosis (β_2)
1.0	1.0	2.0000	9.0000
2.0	1.0	0.6311	3.2451
4.0	1.0	−0.0872	2.7478
0.5	1.0	2.0000	9.0000

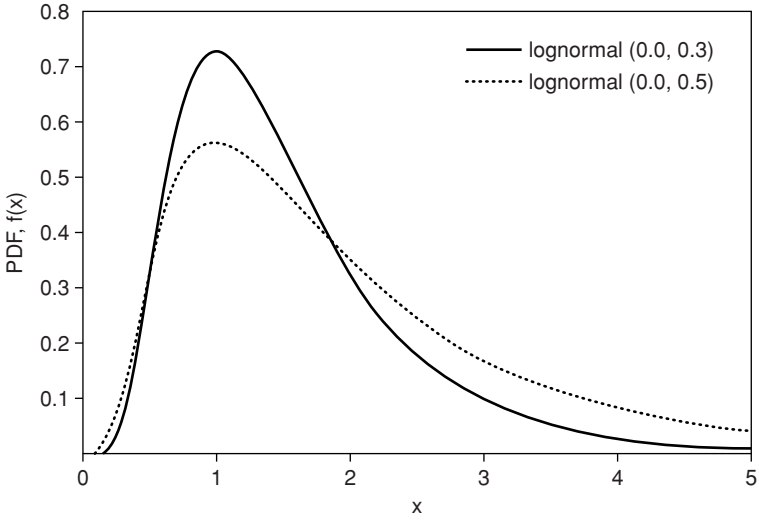


Figure 9.1 PDFs of lognormal distributions with parameters given in Table 9.1.

Each run was replicated 100 times to obtain the average of 100 $\hat{C}_{pu}$ values, $\bar{C}_{pu}$. The simulations results for lognormal and Weibull distributions are given in Tables 9.3–9.6. To investigate the best method for dealing with nonnormality, we present box plots of $\hat{C}_{pu}$ for all seven methods at each targeted C_{pu} and for each distribution (Figures 9.3–9.14). These box plots are able to graphically display several important features of the simulated $\hat{C}_{pu}$ values, such as the mean, variability and outliers. Box plots that are likely to capture the target C_{pu} will have their mean value intersect a horizontal line at the target value.

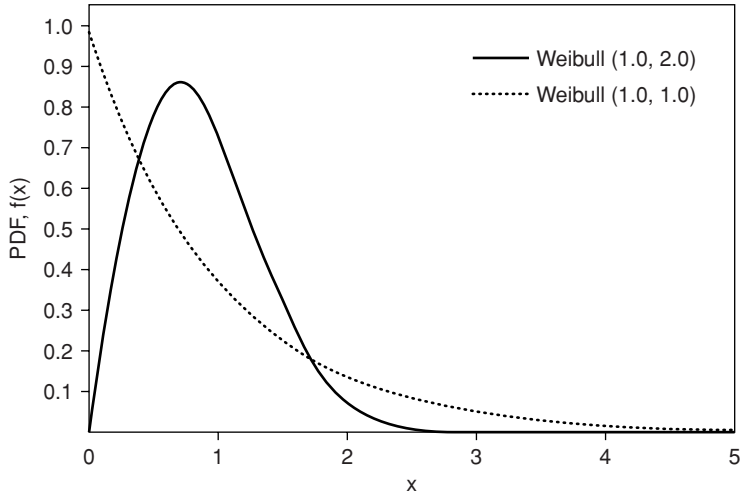


Figure 9.2 PDFs of Weibull distributions with parameters given in Table 9.1.

Table 9.3 Average C_{pu} values calculated from lognormal distribution with $\mu = 0.0$ and $\sigma^2 = 0.5$ ($n = 100$).

C_{pu} (ppm)	*USL*	Assumed normal	Probability plot	Distribution-free tolerance interval	Weighted variance	Clements' method	Box–Cox transform	Johnson transform	Wright's index
1.000 (1349.667)	8.3421	2.3708	1.0146	2.3911	2.1328	1.3082	1.0131	1.1298	1.3619
1.500 (3.401)	24.0947	7.6126	3.1990	7.6586	6.7626	3.3144	1.4693	1.2011	4.3287
1.667 (0.287)	34.1317	10.9608	4.5892	11.7908	9.7368	4.3658	1.6832	1.2591	6.2324

Table 9.4 Average C_{pu} values calculated from lognormal distribution with $\mu = 0.0$ and $\sigma^2 = 0.3$ ($n = 100$).

C_{pu} (ppm)	*USL*	Assumed normal	Probability plot	Distribution-free tolerance interval	Weighted variance	Clements' method	Box–Cox transform	Johnson transform	Wright's index
1.000 (1349.667)	5.175	1.9824	1.0127	1.9408	1.8019	1.1340	1.0067	1.0517	1.2260
1.500 (3.401)	11.7606	5.2363	2.6121	5.1273	4.7593	3.6696	1.5137	1.3812	3.2377
1.667 (0.287)	15.4021	7.0345	3.4960	6.8884	6.3938	4.9186	1.6878	1.4526	4.3495

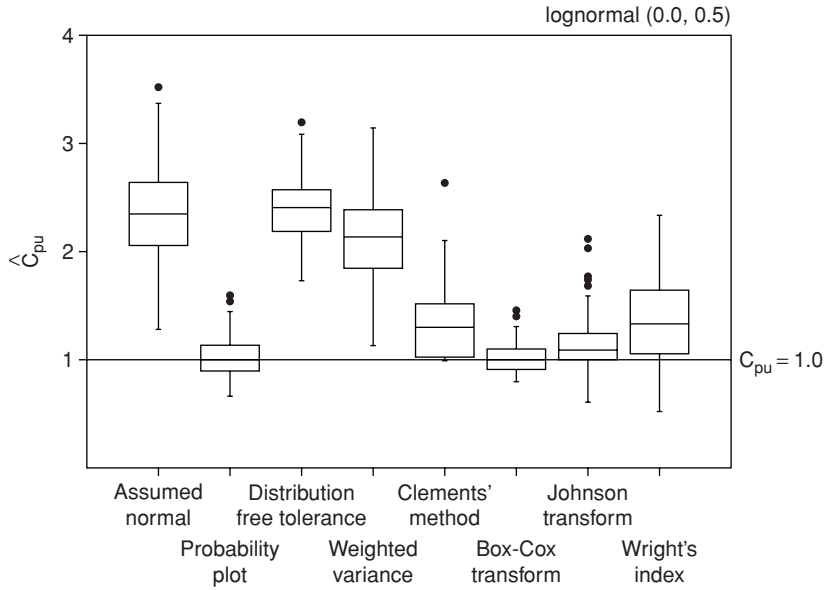


Figure 9.3 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for lognormal ($\mu = 0$, $\sigma^2 = 0.5$) with target $C_{pu} = 1.0$.

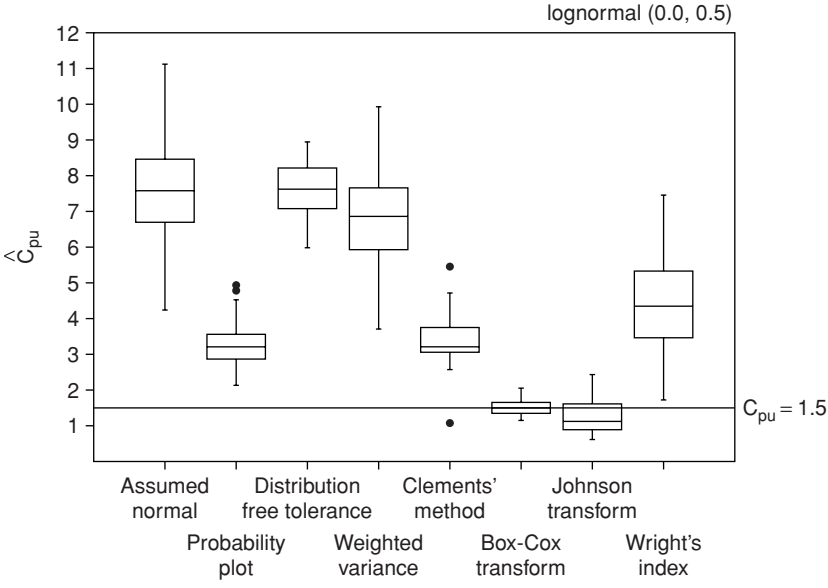


Figure 9.4 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for lognormal ($\mu = 0$, $\sigma^2 = 0.5$) with target $C_{pu} = 1.5$.

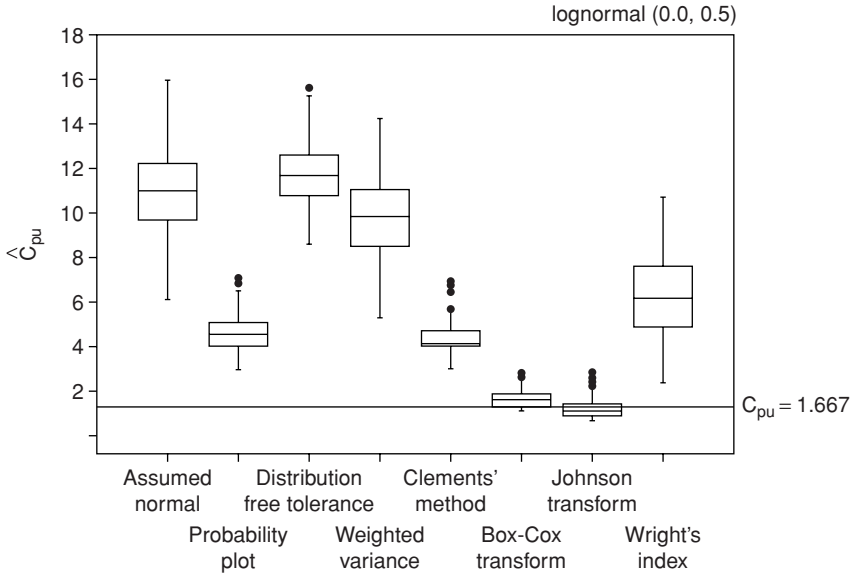


Figure 9.5 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for lognormal ($\mu = 0$, $\sigma^2 = 0.5$) with target $C_{pu} = 1.667$.

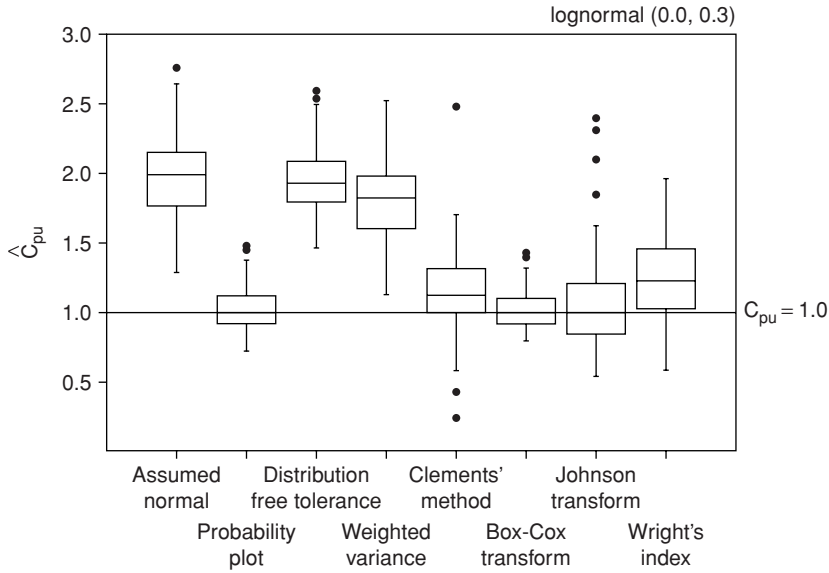


Figure 9.6 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for lognormal ($\mu = 0$, $\sigma^2 = 0.3$) with target $C_{pu} = 1.0$.

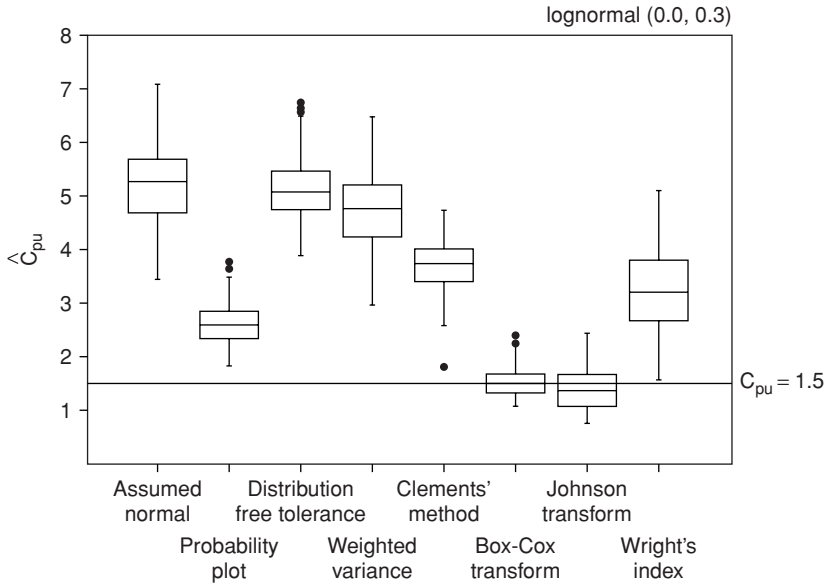


Figure 9.7 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for lognormal ($\mu = 0$, $\sigma^2 = 0.3$) with target $C_{pu} = 1.5$.

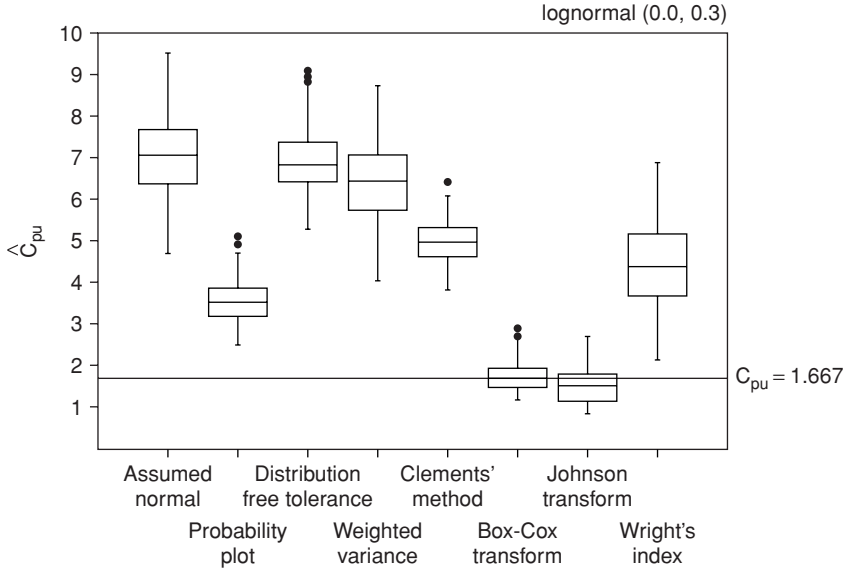


Figure 9.8 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for lognormal ($\mu = 0$, $\sigma^2 = 0.3$) with target $C_{pu} = 1.667$.

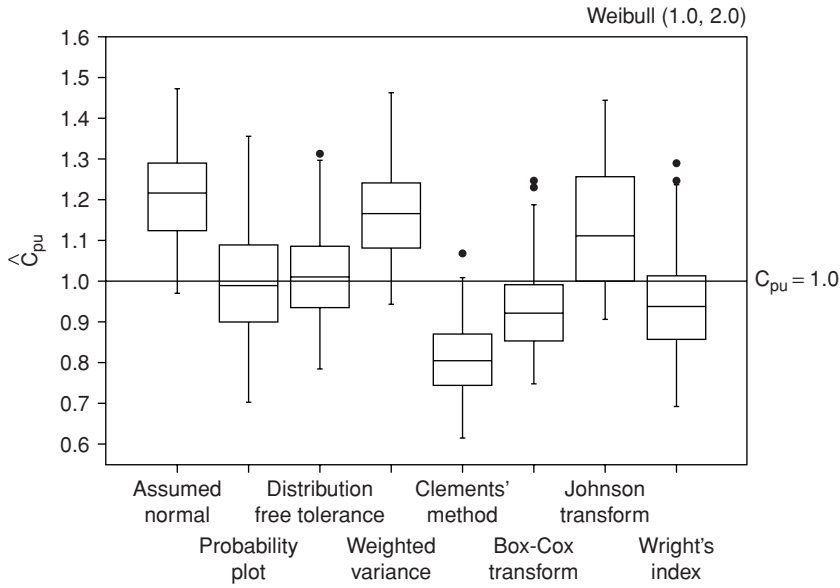


Figure 9.9 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for Weibull ($\mu = 0$, $\eta = 1.0$) with target $C_{pu} = 1.0$.

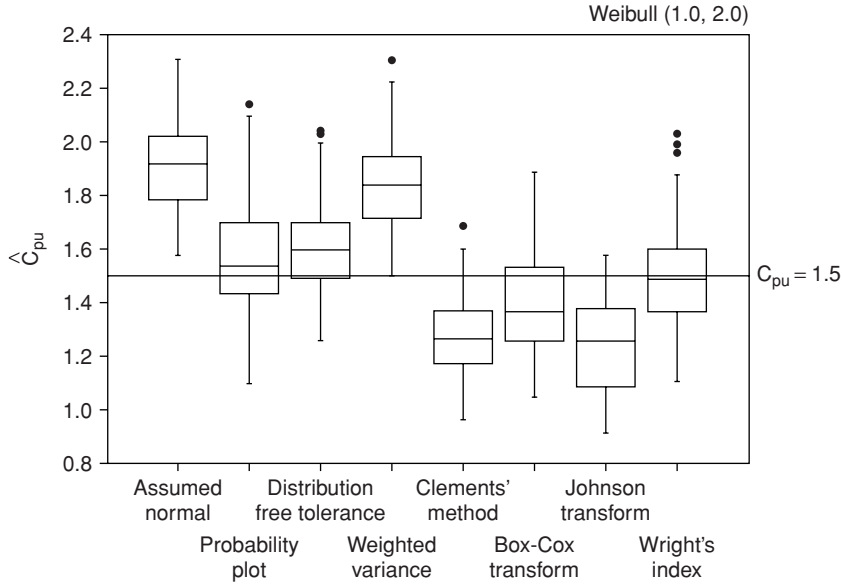


Figure 9.10 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for Weibull ($\sigma = 1.0$, $\eta = 2.0$) with target $C_{pu} = 1.5$.

Table 9.5 Average C_{pu} values calculated from Weibull distribution with $\mu = 1.0$ and $\eta = 2.0$ ($n = 100$).

C_{pu} (ppm)	*USL*	Assumed normal	Probability plot	Distribution-free tolerance interval	Weighted variance	Clements' method	Box–Cox transform	Johnson transform	Wright's index
1.000 (1349.667)	2.5705	1.2121	1.0002	1.0098	1.1622	0.8092	0.9260	1.1233	0.9451
1.500 (3.401)	3.5485	1.9167	1.5647	1.5968	1.8379	1.2722	1.3909	1.2212	1.4949
1.667 (0.287)	3.8812	2.1564	1.7568	1.7965	2.0677	1.4296	1.5333	1.2524	1.6820

Table 9.6 Average C_{pu} values calculated from Weibull distribution with $\mu = 1.0$ and $\eta = 1.0$ ($n = 100$).

C_{pu} (ppm)	*USL*	Assumed normal	Probability plot	Distribution-free tolerance interval	Weighted variance	Clements' method	Box–Cox transform	Johnson transform	Wright's index
1.000 (1349.667)	6.6077	1.8955	1.0058	1.8500	1.6944	1.1479	0.9261	1.0076	1.1581
1.500 (3.401)	12.5915	3.9178	2.0250	3.8236	3.5021	2.2861	1.3056	1.1248	2.3938
1.667 (0.287)	15.0635	4.7532	2.4461	4.6390	4.2489	2.7927	1.4232	1.2542	2.9042

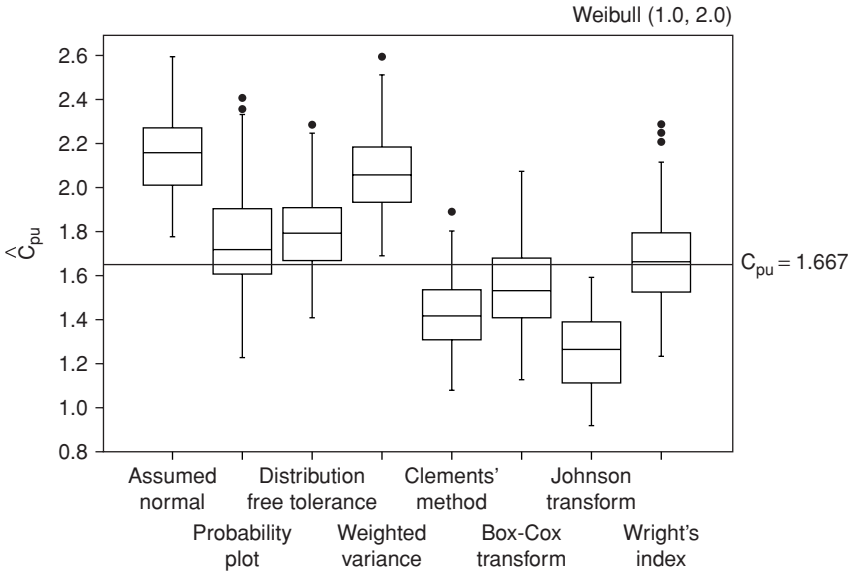


Figure 9.11 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for Weibull ($\sigma = 1.0$, $\eta = 2.0$) with target $C_{pu} = 1.667$.

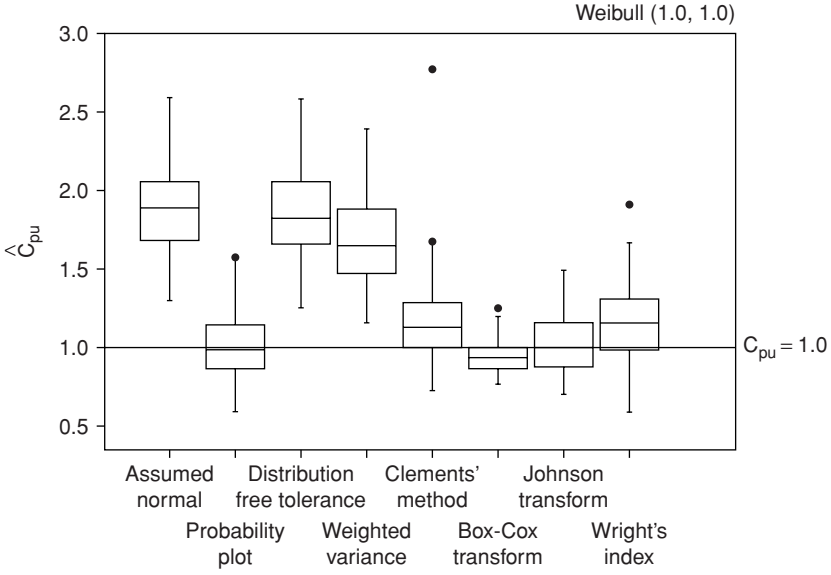


Figure 9.12 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for Weibull ($\sigma = 1.0$, $\eta = 1.0$) with target $C_{pu} = 1.0$.

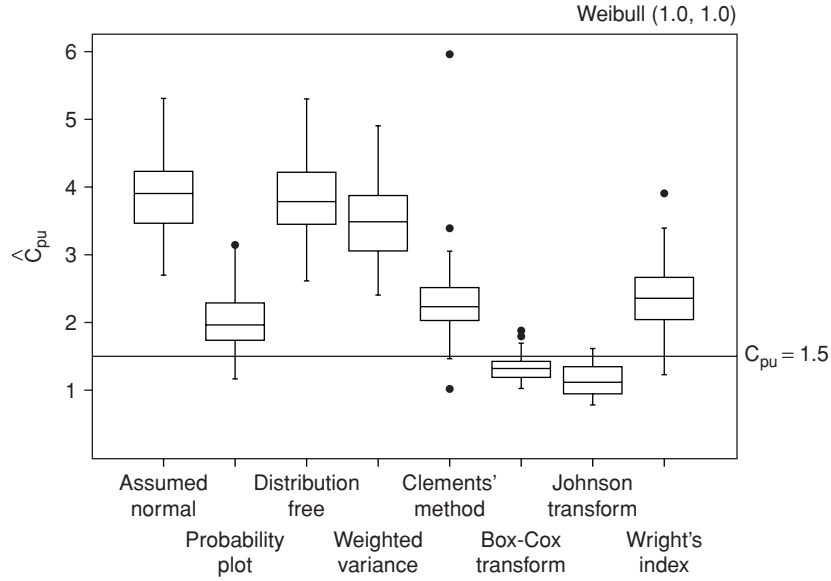


Figure 9.13 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for Weibull ($\sigma = 1.0$, $\eta = 1.0$) with target $C_{pu} = 1.5$.

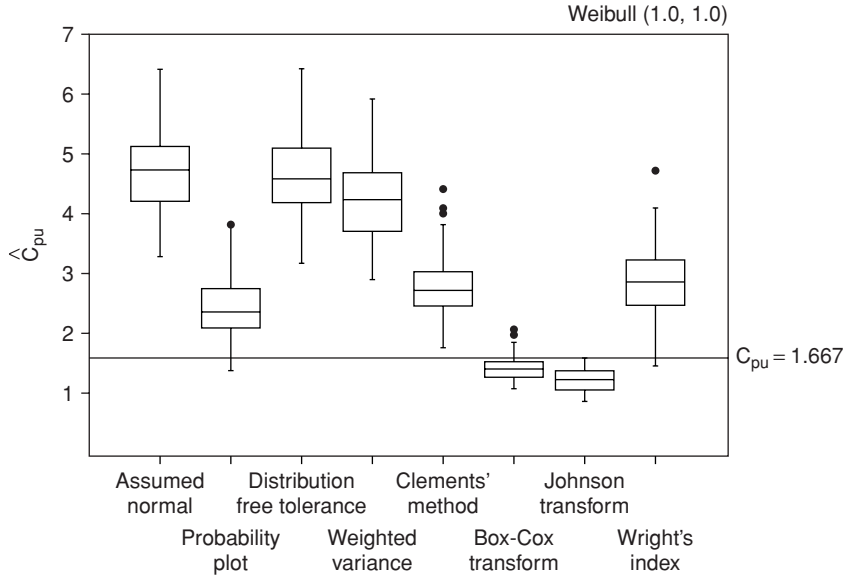


Figure 9.14 Box plots of $\hat{C}_{pu}$ for all methods (100 observations each) for Weibull ($\sigma = 1.0$, $\eta = 1.0$) with target $C_{pu} = 1.667$.

9.4 DISCUSSION OF SIMULATION RESULTS

The seven methods included in this simulation can generally be classified into two categories. The "assumed normal", probability plotting, distribution-free tolerance interval, weighted variance, and Wright's index methods are classified as nontransformation methods. Among the transformation methods are Clements, Box–Cox and Johnson. There are two performance yardsticks in this investigation, namely accuracy and precision given the sample size. For accuracy we look at the difference between the average simulated $\hat{C}_{pu}$ values, $\bar{C}_{pu}$, and the target C_{pu} value. For precision we look at the variation or spread in the average simulated $\hat{C}_{pu}$ values. The smaller the variation or spread in the average simulated $\hat{C}_{pu}$ values, the better is the performance of the method.

From Tables 9.3–9.6 we observe that the performance of the transformation methods is consistently better than that of the non-transformation methods for all the different underlying nonnormal distributions. There are only two exceptions to this statement. Firstly, Clements' method does not perform as well as the other two transformation methods. In the case of the Weibull distribution (see Tables 9.5 and 9.6), the performance of Clements' method is inferior to that of the probability plotting method. Secondly, we observe from the box plots in Figures 9.9–9.11 that the probability plotting method is the only non-transformation method that manages to outperform the transformation methods, but only in the case of the Weibull distribution with $\sigma = 1.0$ and $\eta = 2.0$. For the lognormal distribution the transformation methods consistently yield $\bar{C}_{pu}$ values that are closer to the target C_{pu}, while all the other methods overestimate it. The practical implication is that, though appealing in terms of simplicity of calculation, the other nontransformation methods are found to be inadequate in capturing the capability of the process except when the underlying distribution is close to normal. Moreover, the variations in the simulated $\hat{C}_{pu}$values are consistently large than those of the transformation methods. The use of the other nontransformation methods is clearly inadequate for distributions that depart severely from normal. This means that if one does not recognize the need to work with transformed normal data, one may have a false sense of process capability where the potential process fallout will be greater than expected. Indeed, many practitioners have the experience of getting an acceptable PCI ($\geq$1.33) but suffering from a relatively high process fallout.

Thus far, transformation methods seem to be adequate for handling nonnormal data. The performance of the transformation methods is fairly consistent in terms of accuracy and precision. However, it is noted that the transformation methods are quite sensitive to sample size, as the differences between $\bar{C}_{pu}$ values estimated at $n = 50$ and 150 are 33%, 11%, and 26% for the Clements, Box–Cox, and Johnson transformation methods, respectively. These differences are higher compared with the less than 10% difference in the nontransformation methods. Since, as noted earlier, the transformation methods exhibit a smaller variability in $\hat{C}_{pu}$, such a large difference between $\hat{C}_{pu}$s, computed from small and large sample sizes, cannot be solely attributed to natural statistical variation. This suggests that transformation methods are not appropriate for small sample sizes. Some possible reasons are as follows.

1. Transformation methods usually involve a change from a scale–shape distribution to a location–scale distribution. This may induce an unwanted process shift that distorts $\hat{C}_{pu}$, especially for small sample sizes.

Table 9.7 Practitioners' guide: methods applicable for each defined range of skewness $\sqrt{\beta_1}$ and kurtosis (β_2).

	$-0.1 < \sqrt{\beta_1} \leq 0.1$		$0.1 < \sqrt{\beta_1} \leq 0.1$		$\sqrt{\beta_1} > 0.1$
$2.5 < \beta_2 \leq 3.5$	AN	CM	**PP**	**BC**	**BC**
	PP	**BC**	DF	JT	JT
	DF	JT	WI		
	WV	WI			
$\beta_2 > 3.5$	**BC**		**BC**		**BC**
	JT		JT		JT

PP, probability plot; AN, assumed normal; DF, distribution-free tolerance interval; WV, weighted variance; WI, Wright's index; CM, Clements' method ($n \geq 100$); BC, Box–Cox transformation ($n \geq 100$); JT, Johnson transformation ($n \geq 100$).

2. For the purpose of capability analysis, the success of a transformation method depends on its ability to capture the tail behavior of the distribution (beyond the *USL* percentile point), particularly for high-yield or highly capable processes (see case studies in Chapter 10).

This means that to apply transformation methods in a process capability study, we should have $n \geq 100$.

Nevertheless, the superiority of the Box–Cox transformation is readily seen from the box plots, especially for the lognormal distribution (see Figures 9.3–9.8). This is to be expected, since the Box–Cox transformation method provides an exact transformation by taking the natural logarithms of the lognormal variates to give a normal distribution. Comparing the Box–Cox and Johnson transformation methods, the Box–Cox is more accurate and precise. However, both methods are computer-intensive, with the Box–Cox method requiring maximization of a log-likelihood function to obtain an optimal λ value and the Johnson method requiring fitting of the first four moments to determine the appropriate Johnson family. This leads us to suggest that the Box–Cox transformation is the preferred method for handling nonnormal data whenever a computer-assisted analysis is available.

Combining these results, we present in Table 9.7 a practitioners' guide which shows the methods applicable for each defined range of skewness and kurtosis (see Table 9.2). Note that these ranges are not exact; rather they are inferred from the simulation results available. In Table 9.7, methods that are considered "superior", namely those that give more accurate and precise estimates of C_{pu}, are shown in bold type. Methods that are not highlighted in this way are recommended on the basis of their case of calculation and small bias in estimating the C_{pu} index. It is evident that the Box–Cox transformation is consistently superior in performance throughout. The use of probability plotting is recommended for processes that exhibit a mild departure from normality. Furthermore, a probability plot allows the assessment of the goodness of fit of a particular probability model to a set of data.

9.5 CONCLUSION

Seven methods for computing PCIs have been reviewed and their performance evaluated by Monte Carlo simulation. In general, methods involving transformation,

though more tedious, seem to provide estimates of C_{pu} that truly reflect the capability of the process. In particular, for the lognormal distribution, which has a heavy tail, other methods severely overestimate C_{pu}. This can easily be explained by the fact that, under the normality assumption, C_p and C_{pk} jointly determine the proportion of nonconforming items, p. If the process distribution is not normal, this relation is no longer valid. Thus, unless a proposed method is able to provide a PCI that captures such a relation, the surrogate PCI value will not give an objective view of the real capability of the process in terms of p. The simulation results have supported this, as the performance of a method depends on its capability to capture the behaviour of the tail of the distribution. A method that performs well for a particular distribution may give erroneous results for another distribution with a different tail behavior. The effect of the tail area is quite dramatic (see case studies in Chapter 10), especially for more capable processes.

Though computer-intensive, the Box–Cox transformation method yields highly consistent and accurate $\hat{C}_{pu}$ values for all the distributions investigated with reasonable sample size ($n \geq 100$). This is evident from the simulation results. Today, most process capability studies and analyses are carried out using computers. Therefore the computer-intensive nature of the Box–Cox transformation is no longer a hindrance to practitioners. The accuracy of the Box–Cox transformation method is also robust to departures from normal. Therefore this avoids the trouble of having to search for a suitable method for each distribution encountered in practice. The PCIs calculated using the Box–Cox transformation method will definitely give practitioners a more accurate picture of a process's capability, especially in terms of p. Nevertheless, for each reference under various applications a practitioners' guide is proposed that shows the methods applicable for each defined range of skewness and kurtosis.

REFERENCES

1. Kane, V.E. (1986) Process capability indices. *Journal of Quality Technology*, 18, 41–52.
2. English, J.R. and Taylor, G.D. (1993) Process capability analysis – a robustness study. *International Journal of Product Research*, 31, 1621–1635.
3. Kotz, S. and Johnson, N.L. (1993) *Process Capability Indices*. London: Chapman & Hall.
4. Clements, J.A. (1989) Process capability indices for non-normal calculations. *Quality Progress*, 22, 49–55.
5. Johnson, N.L., Kotz, S. and Pearn, W.L. (1992) Flexible process capability indices. *Pakistan Journal of Statistics*, 10, 23–31.
6. Chan, L.K., Cheng, S.W. and Spiring, F.A. (1992) A graphical technique for process capability. *ASQC Quality Congress Transactions, Dallas*, pp. 268–275.
7. Wright, P.A. (1995) A process capability index sensitive to skewness. *Journal of Statistical Computation and Simulation*, 52, 195–203.
8. Choi, I.S. and Bai, D.S. (1996) Process capability indices for skewed populations. *Proceedings of the 20th International Conference on Computer and Industrial Engineering*, pp. 1211–1214.
9. Pearn, W.L. and Kotz, S. (1994) Application of Clements' method for calculating second and third generation process capability indices for nonnormal Pearsonian populations. *Quality Engineering*, 7, 139–145.
10. Castagliola, P. (1996) Evaluation of non-normal process capability indices using Burr's distributions. *Quality Engineering*, 8, 587–593.
11. Vännman, K. (1995) A unified approach to capability indices. *Statistica Sinica*, 5, 805–820.
12. Deleryd, M. (1996) Process capability indices in theory and practices. Licentiate thesis, Division of Quality Technology and Statistics, Luleå University of Technology.

13. Bissell, D. (1994) *Statistical Methods for SPC and TQM*. London: Chapman & Hall.
14. Gruska, G.F., Mirkhani, K. and Lamberson, L.R. (1989) *Non-normal Data Analysis*. Garden City, MI: Multiface Publishing Co.
15. Box, G.E.P. and Cox, D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society B*, 26, 221–252.
16. Johnson, N.L. (1949) Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149–176.
17. Hahn, G.J. and Shapiro, S.S. (1967) *Statistical Models in Engineering*. New York: John Wiley & Sons, Inc.
18. Hill, I.D., Hill, R. and Holder, R.L. (1976) Fitting Johnson curves by moments algorithm AS99. *Applied Statistics*, 25, 180–189.
19. Tang, L.C., Than, S.E. and Ang, B.W. (1997) A graphical approach to obtaining confidence limits of C_{pk}. *Quality and Reliability Engineering International*, 13, 337–346. See also Chapter 13, this volume.
20. Rivera, L.A.R., Hubele, N.F. and Lawrence, F.D. (1995) C_{pk} index estimation using data transformation. *Computers and Industrial Engineering*, 29, 55–58.

10

Process Capability Analysis for Non-Normal Data with MINITAB

Timothy Yoap

Process capability analysis is a critical component of the Six Sigma methodology. Traditionally, process capability was calculated assuming the data are normally distributed. When the distribution is not normally distributed, it has been widely recognized that the conventional approach in calculating the capability indices is not appropriate. Nevertheless few software package provides routines for performing the analysis. This chapter presents different methods of estimating the process capability for non-normal data using MINITAB.

10.1 INTRODUCTION

Process capability analysis (PCA) is a critical component of Six Sigma, in both the operational (DMAIC) and design (DFSS) spheres. In DMAIC, PCA is normally done during the measure phase to quantify the baseline performance of the process and during the control phase to validate the improvement. In DFSS, the role of PCA is even more important. DFSS starts by understanding the customer's needs and wants. These are then flowed down to the system and eventually the components level. To ensure that the product meets the customer's expectations, each component has to perform at a certain threshold level. To ascertain whether this has been achieved, PCA is needed for each component. The capability score – whether it be z, the number of defects per million opportunities (DPMO) or C_{pk} – for each component will be entered into a management tool called the scorecard. Based on the information in the scorecard, a decision can be made as to whether the product is good enough to be transferred into mass production, and if not, resources can be channeled into work on improving those components that do not meet the capability requirements. Therefore, the accuracy of

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

PCA is of the utmost importance as it directly affects the management's decision-making.

The accuracy of process capability estimation for variable data depends largely on how well the theoretical distribution fits the actual data. In the past, PCA in the industrial world was often done by assuming the process to be normally distributed using the formula

$$C_{pk} = \min\left(\frac{(USL - \bar{x})}{3\sigma}, \frac{(\bar{x} - LSL)}{3\sigma}\right) \tag{10.1}$$

where *USL* and *LSL* denote the upper and lower specification limits, respectively (C_{pk} will be used as the capability measurement in this chapter as it is still the most widely used index today, even though it is not as good a measure as the *z*-score or DPMO). In recent years, however, more and more people have begun to realize that a significant proportion of processes are not normally distributed, leading to serious errors in estimates.

When data is skewed and does not fit the normal distribution, the advice is generally to do a Box–Cox transformation first and use the transformed data to estimate the process capability. This advice is generally good if the optimum power, λ, for the transformation is positive. However, if the λ is negative, it can seriously underestimate the capability of a process. The other approach is to estimate the process capability by using a statistical distribution that provides a good fit to the data.

In this chapter, a case study data set is used to illustrate in detail the two approaches mentioned above. This is followed by another case study data set and three sets of data generated using Monte Carlo simulation with various specification limits. The following steps were followed throughout all the data sets with the aid of the statistical software MINITAB 14.

(a) visual assessment of the process capability by plotting the histogram together with the specification limit;
(b) transformation of the data using the Box–Cox transformation and calculation of the process capability using the formulas above with the transformed data;
(c) fitting the original data with a best-fit statistical distribution and assessing the process capability;
(d) comparison of the result from the different approaches.

As most of the calculations and statistical tests are widely available in statistical software, this chapter focuses on the discussion of the results obtained from the statistical software and not on showing the calculations in detail.

10.2 ILLUSTRATION OF THE TWO METHODOLOGIES USING A CASE STUDY DATA SET

The data set used in this section is taken from a process with mean 12.61 and standard deviation 1.38. The specification limit of this process is 91. From the distribution of the data shown in Figure 10.1, we should expect a C_{pk} much higher than 3 as the

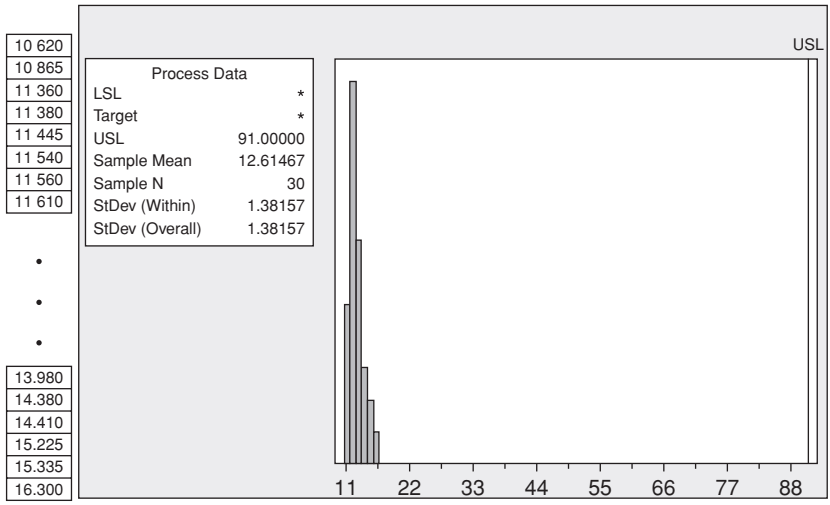


Figure 10.1 Histogram of case study data.

specification limit is so far away from the mean and we do not expect *USL* to be exceeded.

As the distribution is highly skewed, a normal approximation will not yield a good estimate. Two methods were used to estimate the capability of this process; these are considered in turn in Sections 10.2.1 and 10.2.2 and compared in Section 10.2.3.

10.2.1 Estimation of process capability using Box–Cox transformation

Traditionally, when the distribution of the data is not normal, the advice is to first transform the data using the Box–Cox transformation. If the transformation is successful, and the transformed data follows the normal distribution as proven by the normality test, then we can proceed to do a PCA using the transformed data and the transformed specification using formula (10.1).

10.2.1.1 Box–Cox transformation

The Box–Cox transformation is also known as the power transformation. It is done by searching for a power value, λ, which minimizes the standard deviation of a standardized transformed variable. The resulting transformation is $Y' = Y^{\lambda}$ for $\lambda \neq 0$ and $Y' = \ln Y$ when $\lambda = 0$. Normally, after the optimum λ has been determined, it is then rounded either up or down to a number that make some sense; common values are -1, $-1/2$, 0, $1/2$, 2, and 3.

The transformation can be done by most statistical software. Figure 10.2 was obtained from MINITAB 14. It shows a graph of the standard deviation over the different values of λ from -5 to 5. The optimum λ (corresponding to lowest standard deviation) suggested was $-3.537\,25$. It is recommended to choose $\lambda = -3$; it is evident from Figure 10.2 that this gives a standard deviation close to the minimum.

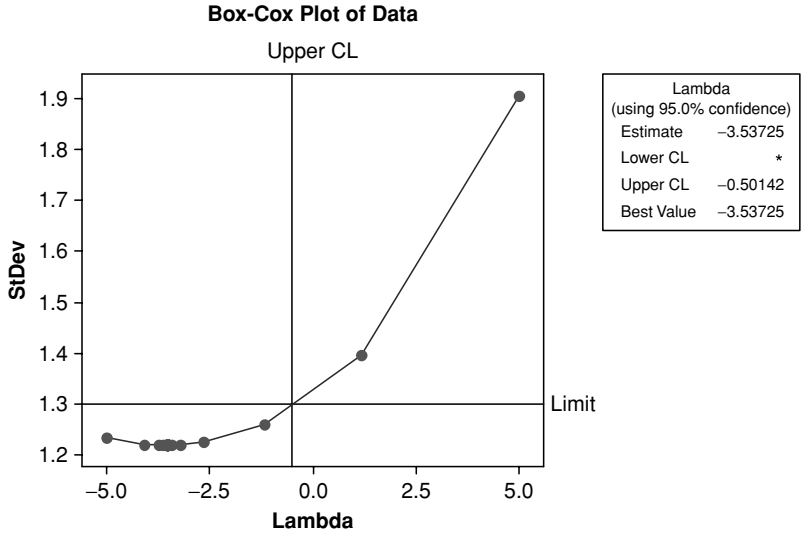


Figure 10.2 Box–Cox plot for case study data.

10.2.1.2 Verifying the transformed data

Having transformed the data, we need to check that the transformation is a good one using both a normality test and visual assessment (fitting a normal curve over the histogram).

There are many normality tests available. The Anderson–Darling test is the test recommended by Stephens.[1] It tests the null hypothesis that the data come from a specified normal distribution by measuring the area between the fitted line based on the distribution and the nonparametric step function based on the plotted points. The statistic is a squared distance and is weighted more heavily in the tails of the distribution. It is given by

$$A^2 = -N - \frac{1}{N}\sum (2i-1)(\ln F(Y_i) + \ln(1 - F(Y_{N+1-i})),$$

where the Y_i are the ordered data and

$$F(Y_i) = \Phi\left(\frac{Y_i - \bar{x}}{s}\right)$$

is the cumulative distribution function of the standard normal distribution. Smaller Anderson–Darling values indicate that there is are fewer differences between the data and the normal distribution, and hence that the data fits the specific distribution better.

Another quantitative measure for reporting the result of the Anderson–Darling normality test is the p-value, representing the probability of concluding that the null hypothesis is false when it is true. If you know A^2 you can calculate the p-value. Let

$$A'^2 = A^2 * \left(1 + \frac{0.75}{N} + \frac{2.25}{N^2}\right),$$

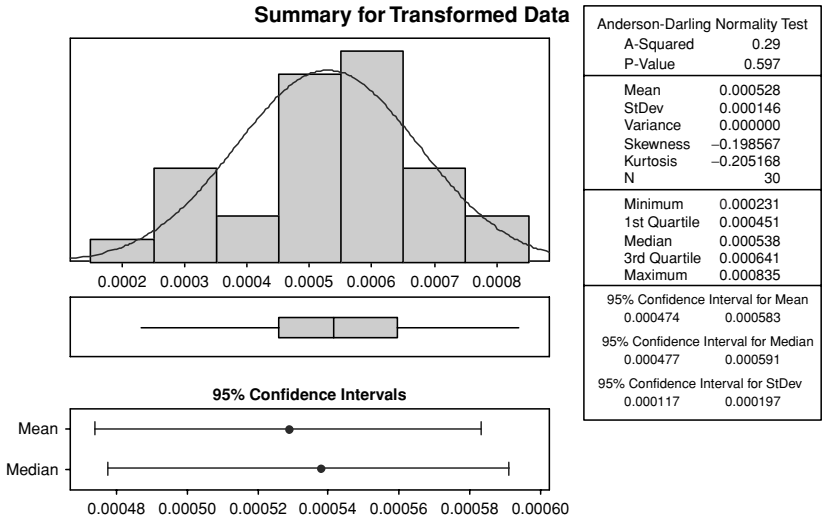


Figure 10.3 Histogram of case study data after Box–Cox transformation.

where N is the sample size. Then

$$p = \begin{cases} \exp(1.2937 - 5.709A'^2 + 0.0186(A'^2)^2), & \text{if } 13 > A'^2 > 0.600 \\ \exp(0.9177 - 4.279A'^2 - 1.38(A'^2)^2), & \text{if } 0.600 > A'^2 > 0.340 \\ 1 - \exp(-8.318 + 42.796A'^2 - 59.938(A'^2)^2), & \text{if } 0.340 > A'^2 > 0.200 \\ 1 - \exp(-13.436 + 101.14A'^2 - 223.73(A'^2)^2), & \text{if } A'^2 < 0.200 \end{cases}$$

Generally if the p-value of the test is greater than 0.05, we do not have enough evidence to reject the null hypothesis (that the data is normally distributed).

Any statistical software can again easily do the above. Figure 10.3 was obtained by MINITAB 14, which provides both the histogram plot with a normal curve on top of it and the Anderson–Darling normality test on the same page. From the Anderson–Darling test we can conclude that the transformed data is normally distributed as the p-value was 0.597, way above the critical value of 0.05.

10.2.1.3 Estimate the process capability using the transformed data

Having successfully transformed the data, we can estimate the process capability using the formula below with the transformed data and the transformed specification, $91^{-3} = 1.327 \times 10^{-6}$ (note that as the power is negative, the original upper specification limit becomes the lower specification limit):

$$C_{pk} = \left(\frac{\bar{x} - LSL}{3\sigma} \right).$$

For our example data set $C_{pk} = 1.19$ (see Figure 10.4).

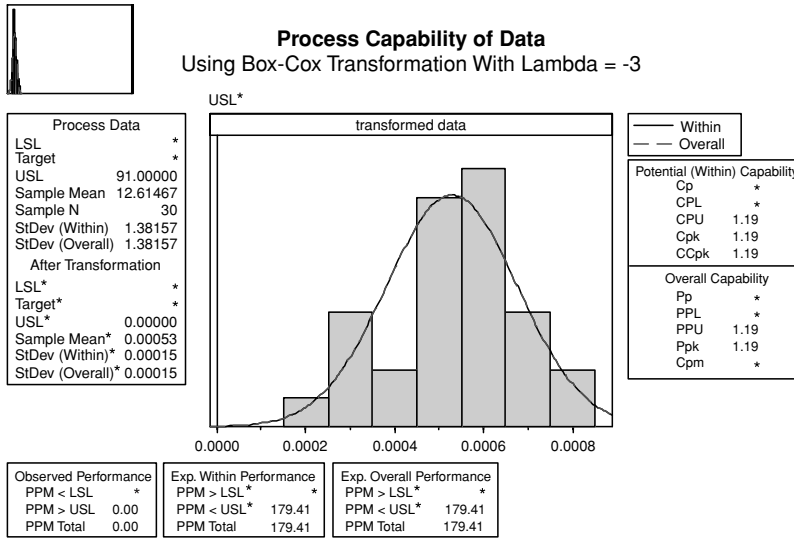


Figure 10.4 Process capability study of case study data using Box–Cox transformation.

10.2.2 Best-fit distribution

An alternative to the Box–Cox transformation is to search for the distribution that best fits the data and use that to estimate the process capability. The search for the best-fit distribution is done by first assuming a few theoretical distributions and then comparing them.

10.2.2.1 Distribution fitting

There are various ways of selecting a distribution for comparison.

1. By the shape of the distribution (skewness and kurtosis). Kurtosis is a measure of how 'sharp' the distribution is as compared to a normal distribution. A high (positive) kurtosis distribution has a sharper peak and fatter tails, while a low (negative) kurtosis distribution has a more rounded peak with wider shoulders.

   Kurtosis can be calculated using the formula

$$\frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^4-\frac{3(N-1)^2}{(N-2)(N-3)}$$

   where x_i is the ith observation, $\bar{x}$ is the mean of the observations, N is the number of nonmissing observations, and s is the standard deviation.

   Skewness is a measure of how symmetrical the distribution is. A negative value indicates skewness to the left (the long tail pointing towards the negative in the pdf), and a positive value indicates skewness to the right (the long tail pointing towards the positive in the pdf). However, a zero value does not necessarily indicate symmetry.

Table 10.1 Common distributions.

Distribution	Parameters	
Smallest extreme value		
Normal	μ = location	$-\infty < \mu < \infty$
Logistic	σ = scale	$\sigma > 0$
Lognormal	μ = location	$\mu > 0$
Loglogistic	σ = scale	$\sigma > 0$
Three-parameter lognormal	μ = location	$\mu > 0$
Three-parameter loglogistic	σ = scale	$\sigma > 0$
	λ = threshold	$-\infty < \lambda < \infty$
Weibull	α = scale	$\alpha = \exp(\mu)$
	β = shape	$\beta = 1/\sigma$
Three-parameter Weibull	α = scale	$\alpha = \exp(\mu)$
	β = shape	$\beta = 1/\sigma$
	λ = threshold	$-\infty < \lambda < \infty$
Exponential	θ = mean	$\theta > 0$
Two-parameter exponential	θ = scale	$\theta > 0$
	λ = threshold	$-\infty < \lambda < \infty$

Skewness can be calculated using the following formula:

$$\frac{N}{(N-1)(N-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

where the symbols have the same meanings as in the kurtosis formula above.

Table 10.1 shows the commonly used distributions and their parameters to be estimated. If the distribution is skewed, try the lognormal, loglogistic, exponential, weibull, and extreme value distributions. If the distribution is not skewed, depending on the kurtosis, try the uniform (if the kurtosis is negative), normal (if the kurtosis is close to zero), Laplace or logistic (if the kurtosis is positive).

2. By nature of the data.

   (a) Cycle time and reliability data typically follow either an exponential or Weibull distribution.
   (b) If the data is screened from a highly incapable process, it is likely to be uniformly distributed.
   (c) If the data is generated from selecting the highest or lowest value of multiple measurements, it is likely to following the extreme value distribution.

10.2.2.2 Parameter estimation

After short-listing the possible distributions that are likely to fit the data well, the parameters for the distribution need to be estimated. The two common estimation methods are least squares and maximum likelihood.

Least squares is a mathematical optimization technique which attempts to find a function which closely approximates the given data. It is done by minimizing the sum of the squares of the error (also known as residuals) between points generated by the function and corresponding points in the data.

Maximum likelihood estimation is a statistical method used to make inferences about parameters of the underlying probability distribution of a given data set. Maximum likelihood estimates of the parameters are calculated by maximizing the likelihood function with respect to the parameters. The likelihood function describes, for each set of distribution parameters, the chance that the true distribution has those parameters based on the sample data.

The Newton–Raphson algorithm can be used to calculate maximum likelihood estimates of the parameters that define the distribution. It is a recursive method for computing the maximum of a function.

10.2.2.3 Selecting the best-fit distribution

Selection of the best-fit distribution can be done either qualitatively (by seeing how well the data points fit the straight line in the probability plot), quantitatively (using goodness-of-fit statistics), or by a combination of the two. Most statistical programs provide the plots and the statistics together.

Probability plot

The probability plot is a graphical technique for assessing whether or not a data set follows a given distribution such as the normal or Weibull.[2] The data are plotted against a theoretical distribution in such a way that the points should approximately form a straight line. Departures from this straight line indicate departures from the specified distribution.

The probability plot provided by MINITAB includes the following:

- plotted points, which are the estimated percentiles for corresponding probabilities of an ordered data set;
- fitted line, which is the expected percentile from the distribution based on maximum likelihood parameter estimates;
- confidence intervals, which are the confidence intervals for the percentiles.

Because the plotted points do not depend on any distribution, they are the same (before being transformed) for any probability plot made. The fitted line, however, differs depending on the parametric distribution chosen. So you can use a probability plot to assess whether a particular distribution fits your data. In general, the closer the points fall to the fitted line, the better the fit.

Anderson–Darling statistic

The Anderson–Darling statistic was mentioned in Section 10.2.1.2. Note that for a given distribution, the Anderson–Darling statistic may be multiplied by a constant (which usually depends on the sample size, n). This is the 'adjusted Anderson–Darling' statistic that MINITAB uses. The p-values are based on the table given by

D'Agostino and Stephens.[1] If no exact p-value is found in the table, MINITAB calculates the p-value by interpolation.

Pearson correlation

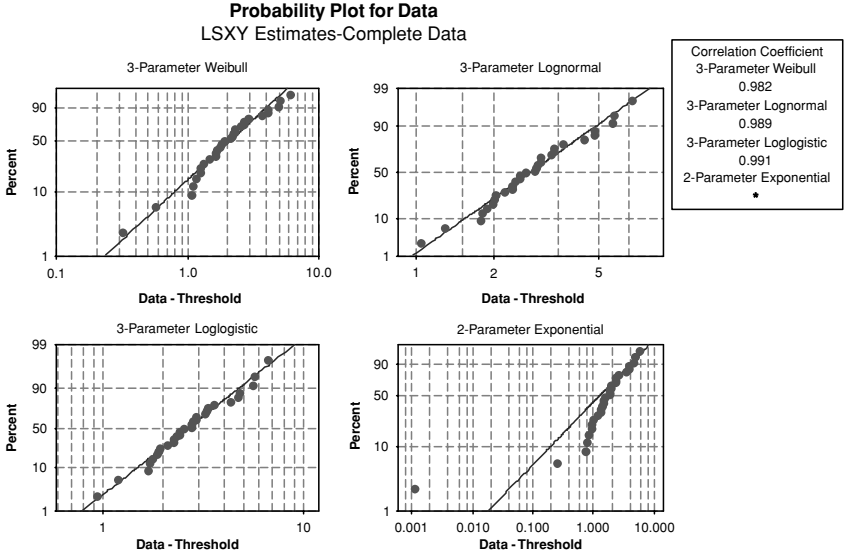
The Pearson correlation measures the strength of the linear relationship between two variables on a probability plot. If the distribution fits the data well, then the plot points on a probability plot will fall on a straight line. The strength of the correlation is measured by

$$r = \frac{\sum z_x z_y}{N}$$

where z_x is the standard normal score for variable X and z_y is the standard normal score for variable Y. If $r = +1(-1)$ there is a perfect positive (negative) correlation between the sample data and the specific distribution; $r = 0$ means there is no correlation.

Capability Estimation for our case study data using the best-fit distribution method

The graph in Figure 10.5 was plotted using MINITAB 14 with two goodness-of-fit statistics, the Anderson–Darling statistic and Pearson correlation coefficient, to help the user compare the fit of the distributions. From the probability plots, it looks like the three-parameter loglogistic and three-parameter lognormal are equally good fits for the data. From the statistics (Anderson–Darling and Pearson correlation), the three-parameter loglogistic is marginally better.



Goodness-of-Fit	Anderson-Darling (adj)	Correlation Coefficient
3-parameter Weibull	1.010	0.982
3-parameter Lognormal	0.698	0.989
3-parameter Loglogistic	0.637	0.991
2-parameter Exponential	3.071	*

Figure 10.5 Probability plots for identification of best-fit distribution.

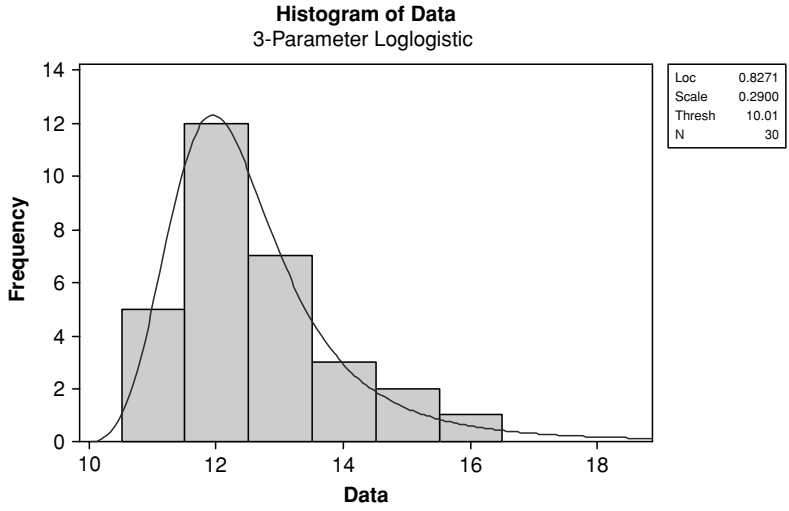


Figure 10.6 Histogram and three-parameter loglogistic fit for case study data.

It is generally good practice to plot the distribution over the histogram make sure the fit is good enough. The graph in Figure 10.6 provides convincing evidence that the three-parameter loglogistic distribution provides an excellent fit for the data.

After the best-fit distribution has been identified, the process capability can be estimated easily by computer. Figure 10.7 shows that the estimated C_{pk} is 5.94, which is close to what we expected. (Note that instead of C_{pk}, P_{pk} is given in the graph. In Six Sigma, P_{pk} is the long-term capability while C_{pk} is the short-term process capability. MINITAB assumes that the data is long term for non-normal data. In this chapter, we will skip the discussion on whether the data is long term or short term, and treat all estimated capability the C_{pk}.)

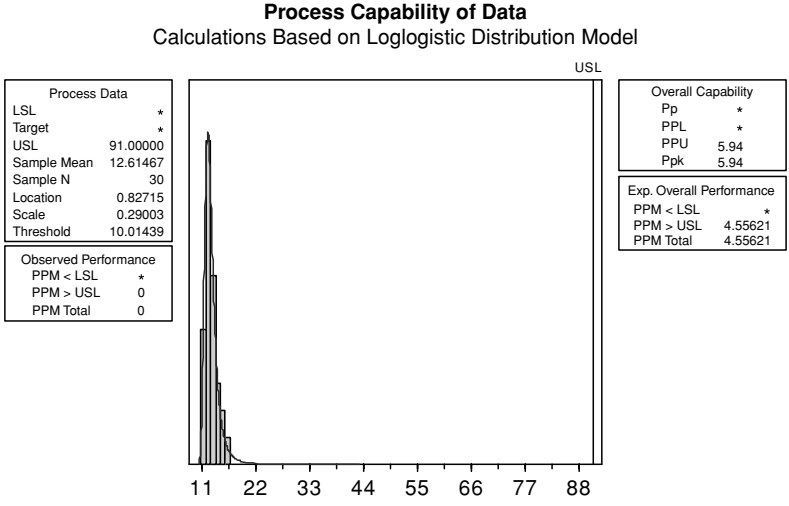


Figure 10.7 Process capability analysis using three-parameter loglogistic fit.

10.2.3 Comparison of Results

The estimation of the process capability, C_{pk}, is 1.19 using the Box–Cox transformation and 5.94 using the best-fit distribution. The difference is 4.75, which is huge. Of course, all estimates are wrong, but some are more wrong than others. What is needed is an estimate that is good enough for decision-making. On which process needs a 100 % screening and which one just needs sampling for Xbar-R chart.

Looking at the data and specification limit for the case study, there is no room for doubt that this is a highly capable process, much more so than the common industrial standard for the C_{pk} (1.33). Whether the C_{pk} is 3 or 5 does not really matter as it does not affect the decision. But using the Box–Cox transformation method alone would lead us to suspect that the process is not capable enough. Relying on this method alone would here led us to commit resources that ultimately would have been wasted.

The reason for the gross inaccuracy when the λ is negative lies in the fact that the minimum value for the transformed data is 0, violating the assumption that normal data can take any value from $-\infty$ to $+\infty$. This violation is not a big problem if the mean of the transformed distribution is far away from zero. But if the mean is close to zero (which in most cases it will be when the distribution is transformed with a negative λ, as any value greater than one will be between zero and one after the transformation), the estimation will think that the distribution has a 'tail' all the way to $-\infty$, and this will bring down the C_{pk} estimate.

10.3 A FURTHER CASE STUDY

Another data set was taken from another process. The summary statistics and the histogram are shown in Figure 10.8. The upper specification limit for this process is 75, and there is no lower specification limit.

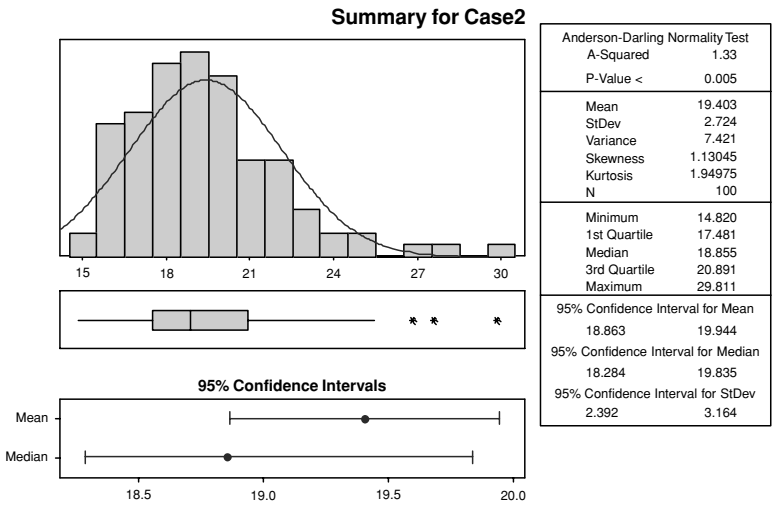


Figure 10.8 Summary statistics for Case Study 2.

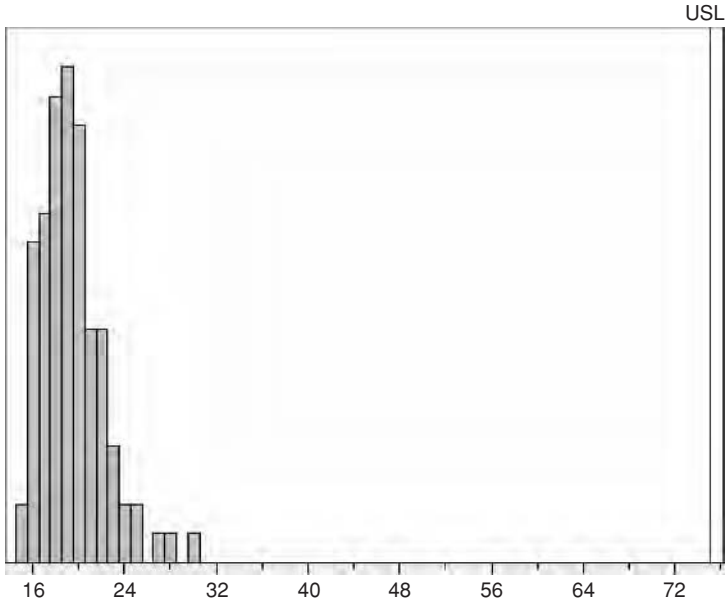


Figure 10.9 Distribution with *USL* for Case Study 2.

The histogram shows that the distribution is skewed and therefore not normally distributed. The Anderson–darling normality test, with p-value way below 0.05, confirms this.

Let us try plotting the distribution together with *USL* to get a feel of the capability of the process (Figure 10.9). As *USL* is more than 15 standard deviations away from the specification limit, we would expect a very high C_{pk} for this process.

We will now try to estimate the C_{pk} of this non-normal distribution using the two approaches introduced in Section 10.2, the Box–Cox transformation method and the estimation using the best-fit statistical distribution.

10.3.1 Process capability analysis using the Box–Cox transformation

The Box–Cox plot in Figure 10.10 shows that the optimum λ is $-1.705\,51$. We may want to round it off to -2, and transform the data by $Y' = Y^{-2}$. The histogram plot and Anderson–Darling test (p-value $= 0.982$) in Figure 10.11 show no evidence to reject the claim that the transformed data is normally distributed. Therefore, we can estimate the C_{pk} by treating the transformed data as normal.

The estimation of the process capability will be done with the transformed value of *USL*, $75^{-2} = 0.000\,178$. (Note that when the power of the Box–Cox transformation is negative, the transformed *USL* will become the *LSL*).

The result of the PCA is shown in Figure 10.12. The C_{pk} is 1.25.

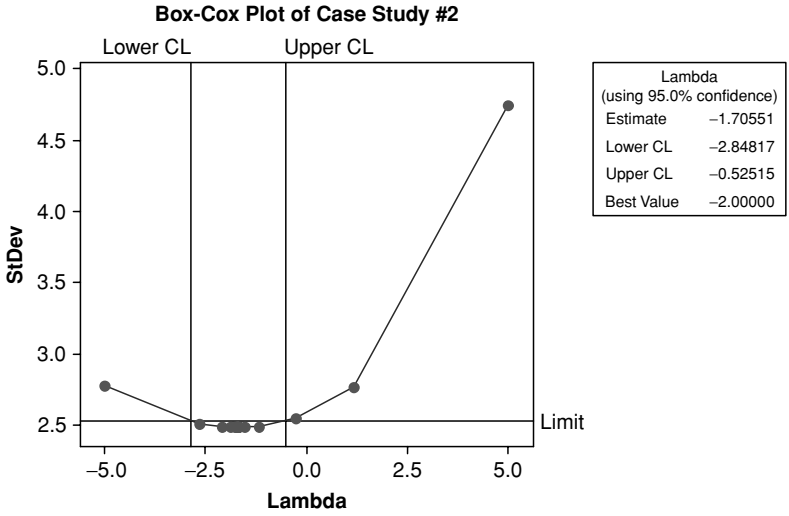


Figure 10.10 Box–Cox plot of Case Study 2.

10.3.2 Process capability analysis using the best-fit distribution

From the probability plot and the Anderson–Darling statistics in Figure 10.13, the best-fit distribution for the data is the three-parameter lognormal distribution.

Using the three-parameter lognormal distribution to fit the data, with $USL = 75$, we obtained the histogram plot with the fitted distribution and the estimated C_{pk} shown

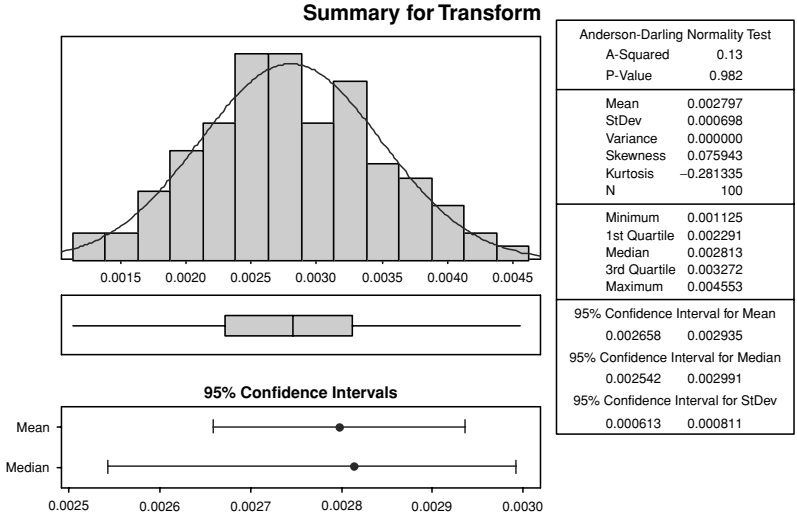


Figure 10.11 Summary statistics and histogram of the transformed data of Case Study 2.

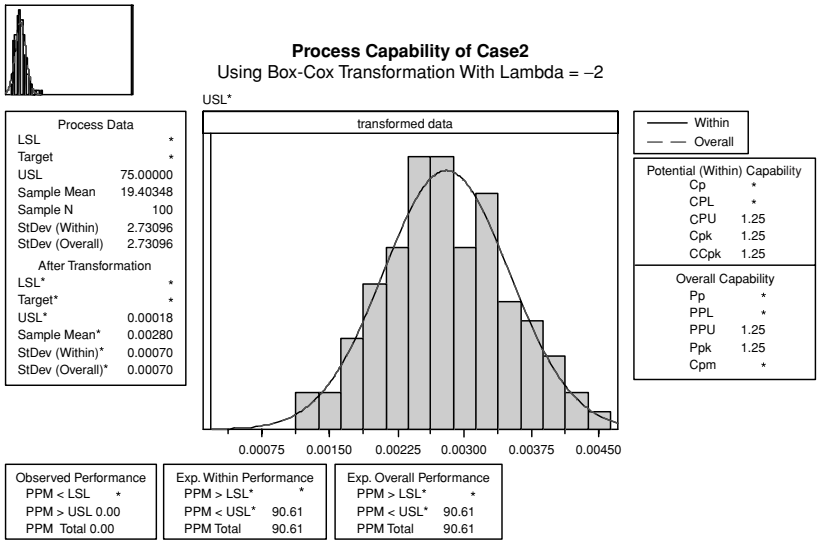
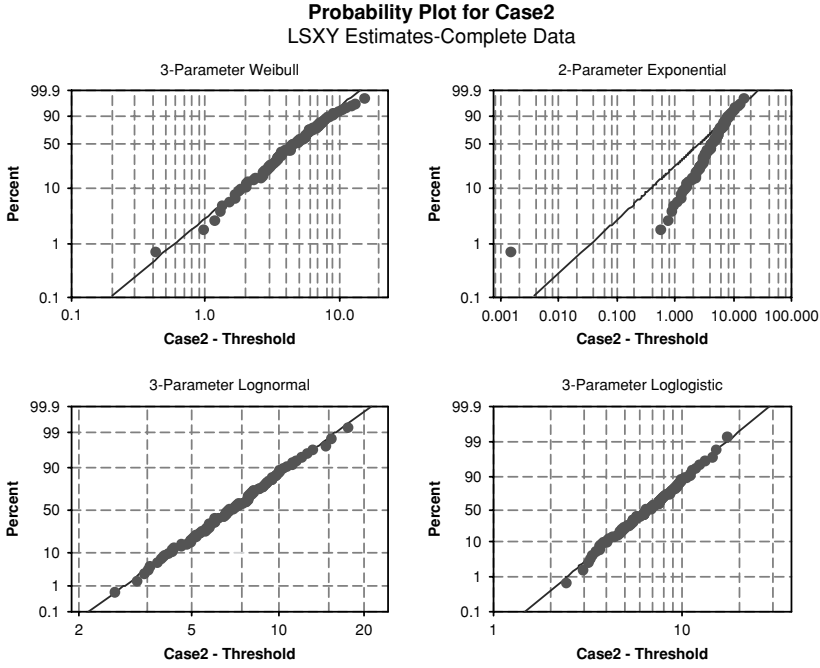


Figure 10.12 Process capability of Case Study 2 using Box–Cox transformation method.



Goodness-of-Fit

Distribution	Anderson-Darling (adj)	Correlation Coefficient
3-Parameter Weibull	0.480	0.995
2-Parameter Exponential	14.282	*
3-Parameter Lognormal	0.237	0.999
3-Parameter Loglogistic	0.295	0.998

Figure 10.13 Probability plot and goodness-of-fit statistics.

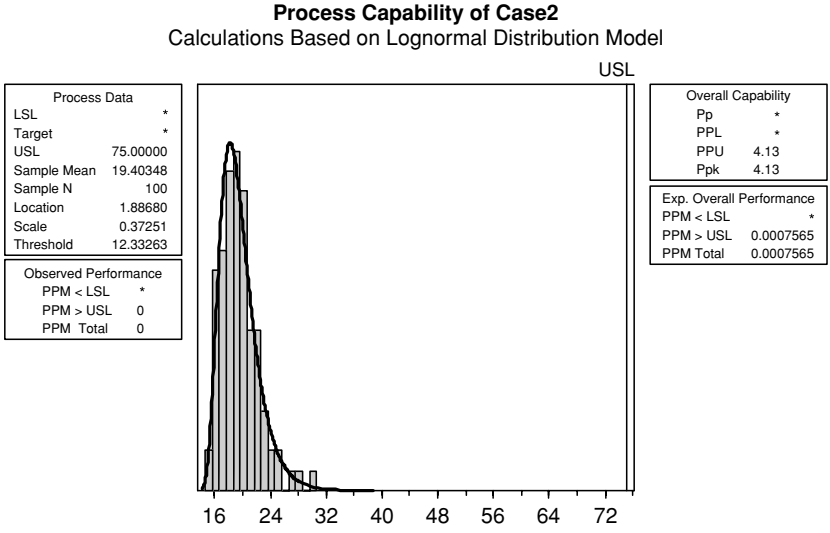


Figure 10.14 Process capability study with three-parameter lognormal fit.

in Figure 10.14. We can see that the three-parameter lognormal distribution actually fitted the data very well. The estimated C_{pk} using this method was 4.13.

10.3.3 Comparison of results

The C_{pk} estimated using Box–Cox transformation method was 1.25, which was 2.88 lower than that estimated by the best-fit distribution. From a Six Sigma decision point of view, a C_{pk} of 1.25 is not good enough and we would need to spend resources looking into it and improving it, whereas a C_{pk} of 4.13 is so good that we should just leave it alone. From Figure 10.14 it is hard to believe that the *USL* of 75 will be exceeded if there is no great change in the process as the specification limit is so far (many standard deviations) away the data concentration. Therefore, from the visual method and the reasoning given in Section 10.2.3, although both methods of estimation are statistically acceptable, the method of using the best-fit distribution is recommended.

10.4 MONTE CARLO SIMULATION

To further study the problem, three sets of data were artificially generated using Monte Carlo simulation with a three-parameter lognormal distribution. The parameters were carefully chosen so that they have an optimum λ of -3, -2 and -1 for Box–Cox transformation. The process capability of each of the data sets was estimated using both the best-fit distribution and normal approximation after Box–Cox transformation. The Upper specification limits between 1 and 20 standard deviations away from the mean were used to understand the differences in C_{pk} estimation with respect to the capability of a process.

10.4.1 Simulated data sets

Data set 1 was generated using the three-parameter lognormal distribution with the following parameters: location = −3.5; scale = 0.2; threshold = 0.1. The histogram with the best-fit distribution and Box–Cox transformed (with $\lambda = -3$) distribution are shown in Figure 10.15. The p-value from the Anderson–Darling normality test is 0.91, indicating we cannot reject the claim that the transformed distribution is normal and the Box–Cox transformation is successful.

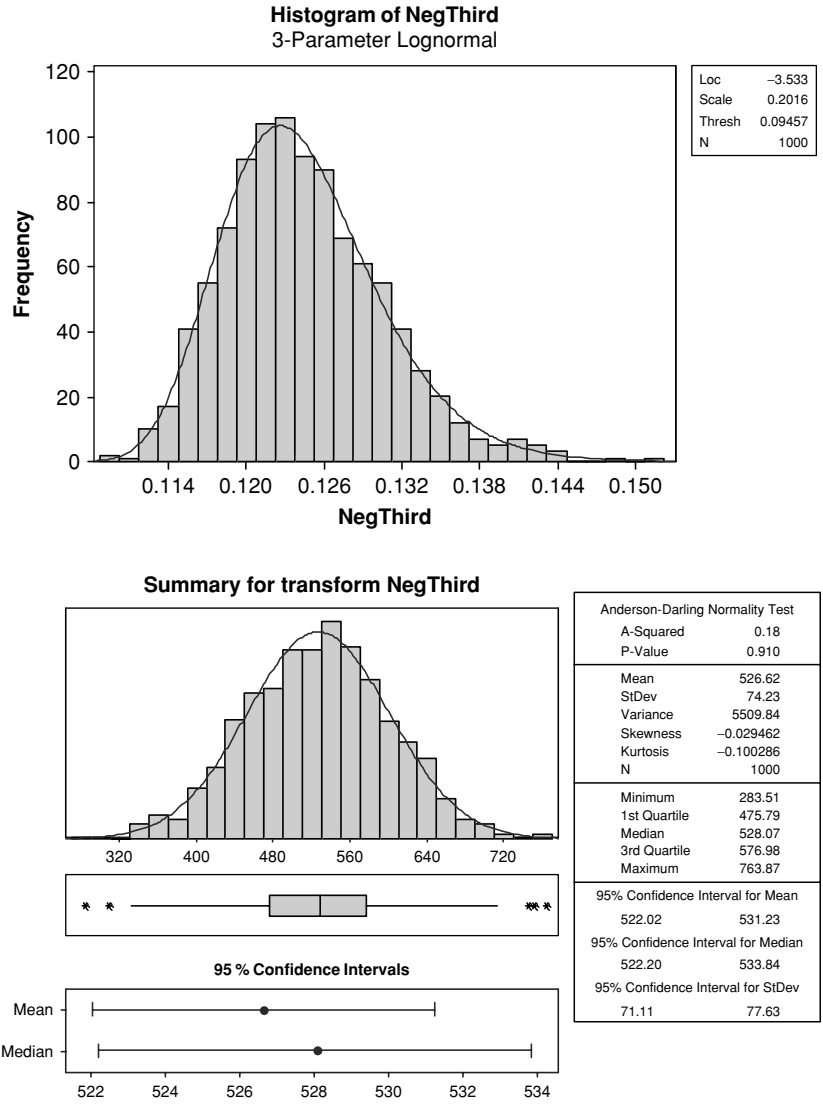


Figure 10.15 Comparison of process capability between Box–Cox with power −3 and lognormal fit.

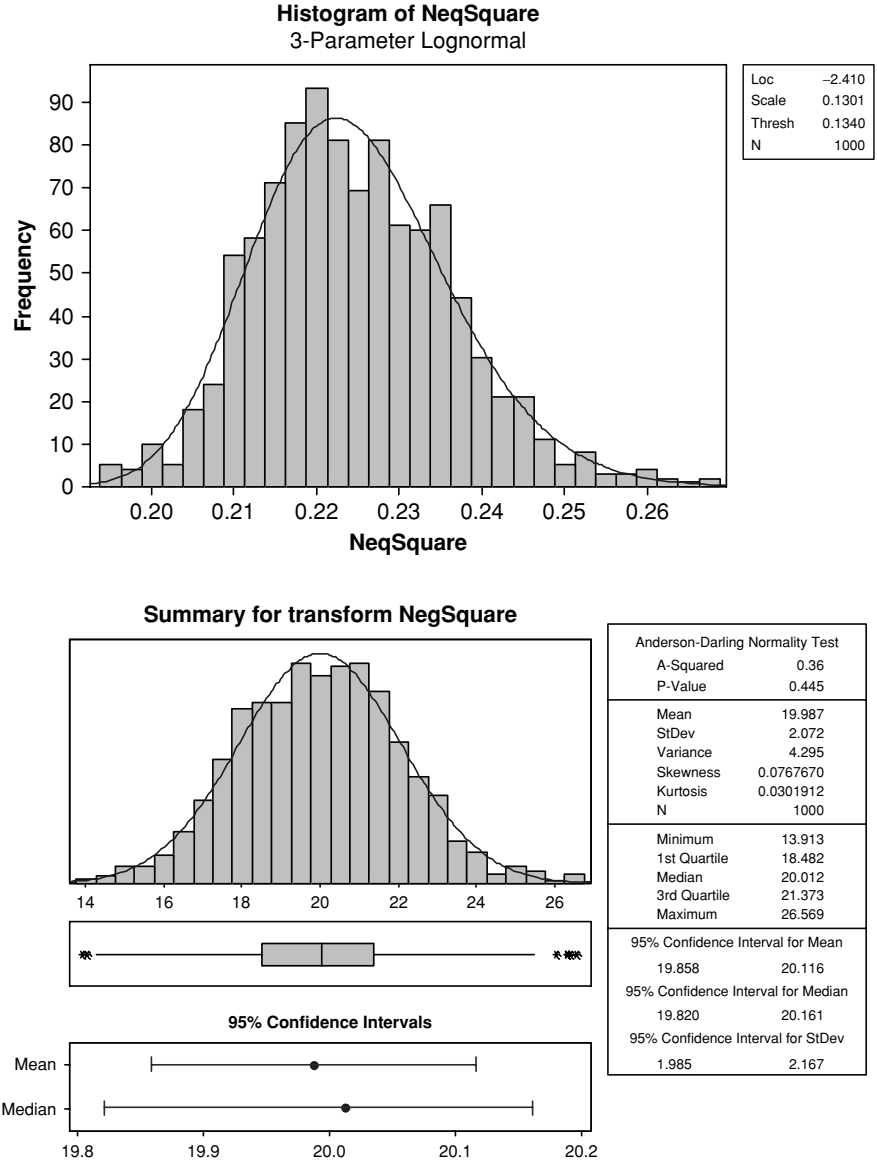


Figure 10.16 Comparison of process capability between Box–Cox with power −2 and three-parameter lognormal fit.

Data set 2 was generated using the three-parameter lognormal distribution with location −2.5, scale 0.13, and threshold 0.1. The histogram with the best-fit distribution and Box–Cox transformed (with $\lambda = -2$) distribution are shown in Figure 10.16. The p-value from the Anderson–Darling normality test is 0.445, indicating we cannot reject the claim that the transformed distribution is normal and the Box–Cox transformation is successful.

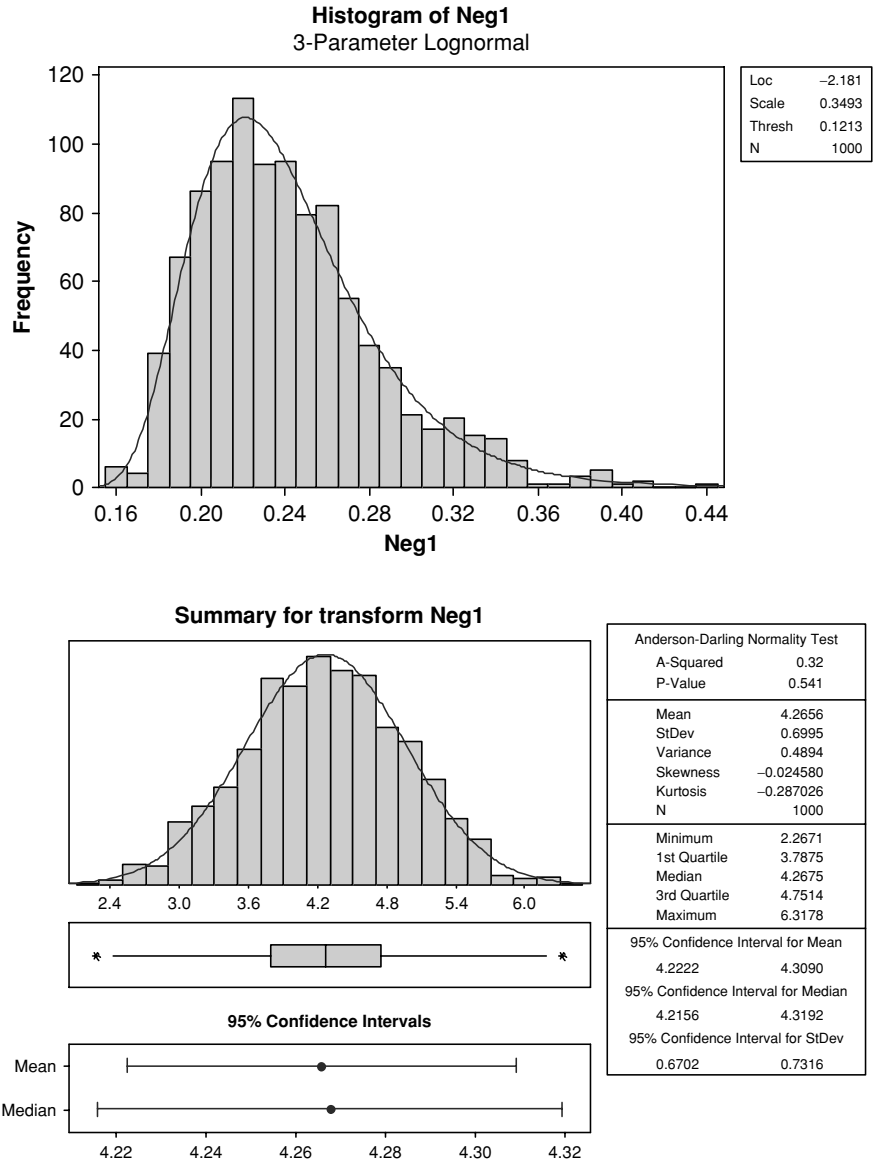


Figure 10.17 Comparison of process capability between Box–Cox with power −1 and lognormal fit.

Data set 3 was generated using the three-parameter lognormal distribution with location −2, scale 0.3, and threshold 0.1. The histogram with the best-fit distribution and Box–Cox transformed (with $\lambda = -1$) distribution are shown in Figure 10.17. The p-value from the Anderson–Darling normality test is 0.541, indicating we cannot reject the claim that the transformed distribution is normal and the Box–Cox transformation is successful.

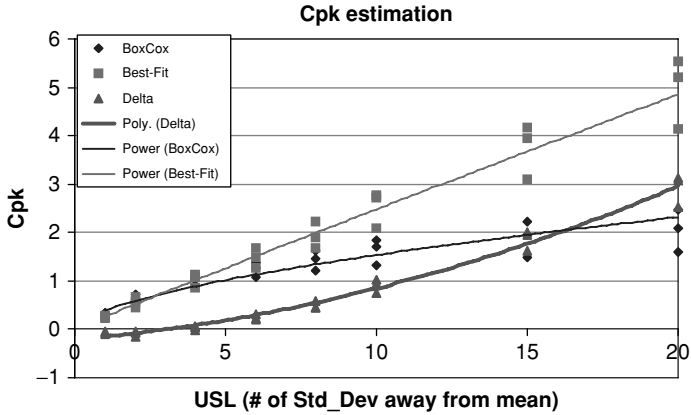


Figure 10.18 Comparison of process capability between Box-Cox with power of −1 vs. Log-normal fit.

10.4.2 Comparison of results

It can be seen from Figure 10.18 that the difference in the estimation of C_{pk} widens as the process grows more capable. When $C_{pk} < 1$, the difference is negligible. When the process capability (as estimated by the Box–Cox method) is more than 2, although the difference is very big, this does not really matter in real life as it does not affect the decision, that is the process is very capable and therefore no effort is required to improve it. The critical region is between them where the difference can change the decision from 'leave it alone' to 'need to improve the process'.

10.5 SUMMARY

When Box–Cox transformation is used in estimation of C_{pk} for a distribution that is skewed to the right, as compared to a reasonable fit using the best-fit distribution, it is likely to be characterized by underestimation. The higher the process capability, the worse the underestimation. When the C_{pk} is low (below 1), estimation using both methods is very close. Therefore, it is advisable not to use Box–Cox transformation for the purpose of process capability analysis when λ is negative.

Today, most statistical software (e.g. MINITAB 14) allows user to select the best-fit distribution for process capability analysis. This makes PCA for non-normal data an easy task without using the Box–Cox transformation.

REFERENCES

1. D'Agostino, R.B. and Stephens, M.A. (1986) *Goodness-of-Fit Techniques*. New York: Marcel Dekker.
2. Chambers, J., Cleveland, W., Kleiner, B. and Tukey, P. (1983) *Graphical Methods for Data Analysis*. Belmont, CA: Wadsworth.

Part C

Analyze Phase

11

Goodness-of-Fit Tests for Normality

L. C. Tang and S. W. Lam

One of the basic model assumptions in statistical procedures used in Six Sigma applications is that of normality of data. Statistical 'goodness-of-fit' (GOF) techniques have been developed in the past decades to assess the adequacy of using the normal distribution to model real-world data so as to limit the risk against severe departures from normality. These GOF techniques have been widely used in Six Sigma projects with the help of statistical software packages such as JMP and MINITAB. This chapter presents the theoretical concepts together with the operational procedures (coded in Microsoft Excel) for a collection of commonly used GOF tests for normality. The limitations of some of these GOF tests are discussed and compared.

11.1 INTRODUCTION

A systematic and highly disciplined data-driven approach which leverages on rigorous statistical analysis procedures for continuous quality improvements has always been the cornerstone for success in Six Sigma projects.[1,2] Only through carefully designed sampling procedures and the rigorous assessment of information can breakthrough Six Sigma improvement results be achieved and objectively justified. Statistical techniques which are built on firm scientific thinking and mathematical foundations provide convenient tools to achieve this in Six Sigma projects.

Central to the practical application of statistical procedures is the statistical modeling of data obtained from real-world processes. For Six Sigma projects to effectively leverage on the strengths of rigorous statistical procedures for quality improvement, the appropriate depiction of real-world data through statistical models is critical before all other data-analysis and optimization procedures can be undertaken. Appropriate

Six Sigma: Advance Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

statistical models can range from theoretical parametric distributional models to more empirical nonparametric or distribution-free models. Apart from the nature of the processes which generate the data, the 'appropriateness' of a statistical model is inadvertently also a function of the data collection and analysis processes. The decision on whether a statistical model is appropriate for a particular data set is typically underpinned by the three fundamental considerations of alignment with theoretical process assumptions, robustness to departures from these assumptions, and downstream data-analysis procedures. On top of these considerations, the model has to be judged on how well it represents the actual data. In order to achieve this, an entire class of statistical techniques known as 'goodness-of-fit' (GOF) tests has been developed. Some of the more popular GOF techniques for assessing the adequacy of the normal distribution in representing the data are reviewed in this chapter. Such GOF tests of normality are commonly encountered in Six Sigma applications as many Six Sigma statistical techniques rely on the normality assumption.

The fundamental statistical hypothesis testing concepts underlying GOF tests are discussed in Section 11.2 as a precursor to setting the correct framework for appropriate applications of these tests. This is followed by a discussion of several popular GOF tests. The basic concepts are presented together with the pros and cons associated with each of these tests. In order to aid understanding, the application procedure is discussed through practical examples for all these tests.

11.2 UNDERLYING PRINCIPLES OF GOODNESS-OF-FIT TESTS

GOF tests were developed primarily from fundamental concepts in statistical hypothesis testing attributed to Neyman and Pearson.[3] In statistical hypothesis testing, there is always a statement of a 'null' hypothesis and an 'alternative' hypothesis which are sets of mutually exclusive possibilities in a sample space. For a typical statistical GOF test these are defined as follows:

$$H_0 : F(x) = F_0(x) \quad \text{vs.} \quad H_1 : F(x) \neq F_0(x)$$

where $F_0(x)$ is some hypothesized distribution function. In Six Sigma applications this is usually taken to be the normal distribution. To this end, it must be stressed that our basic intent here is to limit our risk against severe departure from normality. Generally, the primary aim is *not* in claiming that a particular hypothesized model, represented by the distribution function, $F_0(x)$, is *proven* to be representative of the real data, but to warn us of significant departure from $F_0(x)$. It should also be noted that as "all models are wrong", any pre-conceived $F_0(x)$ is always open to rejection as more information becomes available (i.e. as sample size increases).

GOF tests fall naturally into four broad categories: (1) methods based on discrete classification of data (Pearson χ^2_P); (2) empirical distribution function (EDF) based methods; (3) regression based methods; and (4) methods based on sample moments. Tests of each type are discussed in Sections 11.3–11.7. An example is used to demonstrate the application of each test. Finally, the power of these tests for the data set in the example is compared.

11.3 PEARSON CHI-SQUARE TEST

The Pearson chi-square (χ^2_P) GOF test belongs to a generic family of GOF tests which are based on the X^2 test statistic. The classical X^2 statistic is essentially given by [4,5]

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \tag{11.1}$$

where E_i is the expected frequency of sample observations in class i given that the frequency of samples in each class follows the hypothesized distribution, and O_i is the observed frequency of sample observations in class i. This test requires the classification of random sample observations from a population into k mutually exclusive and exhaustive classes where the number of sample observations in each class is sufficiently large for the expected frequencies (based on some postulated distribution) in each category to be non-trivial. Given that the sample size is n, E_i is given by

$$E_i = np_{i,0},$$

where $p_{i,0}$ is the probability that a sample observation belongs to class i under the null hypothesis

When the null hypothesis holds, the limiting distribution of the X^2 statistic is χ^2 distributed with $(k - c - 1)$ degrees of freedom (where k is the number of non-empty cells and c is the number of parameters to be estimated).[5,6] Given existing proofs based on asymptotic characteristics, it should be noted that the limiting χ^2distribution is a poor approximation to the distribution of X^2 when the number of samples is small. The lack of sensitivity in χ^2 GOF tests with few observations has been frequently acknowledged.[7]

The Pearson χ^2 GOF test is perhaps one of the most commonly used tests due to its versatility. It can easily be applied to test any distributional assumption of discrete or continuous type for any univariate data set without having to know the value of the distributional parameters. The main disadvantage associated with this GOF test is its lack of sensitivity in detecting inadequate models when few observations are available as sufficiently large sample sizes are required for the χ^2assumption to be valid. Furthermore, there is a need to rearrange the data into arbitrary cells. Such grouping of data can affect the outcome of tests as the value of the X^2statistic depends on how data is grouped. Nonetheless, most reasonable choices for grouping the data should produce similar results. Some rule-of-thumb criteria are discussed in the following worked example.

Table 11.1 lists residuals derived from the flight times of model helicopters constructed during a helicopter experiment conducted in a design of experiments experiential learning segment of one of a Design for Six Sigma course conducted by the authors. Such helicopter experiments are described in Box and Liu.[8] Residuals are the differences between the sample observations and the fitted response values obtained from the prediction model. The prediction model is a multiple linear regression model that was established through a factorial experiment. Each residual (e_i) thus describes

Table 11.1 Residuals from prediction model of flight times.

0.06	−0.28	0.54	0.29	−0.33	0.05	−0.19	0.55
0.39	−0.47	−0.39	0.69	−0.19	−0.62	−0.07	−0.58
0.29	0.30	0.15	0.48	−0.08	−0.33	0.24	−0.52
0.32	−0.49	0.07	0.50	−0.04	0.33	0.18	−0.20
0.14	−0.24	0.19	−0.19	0.03	−0.14	0.07	−0.57

the error in the fit of the prediction model ($\hat{Y}_i$) to the ith observation (Y_i) and is given by

$$e_i = Y_i - \hat{Y}_i.$$

In performing the regression analysis the usual assumptions that the errors are independent and follow a normal distribution with zero mean and constant variance are made. Hence, if the prediction model adequately describes the behavior of the actual system, its residuals should not exhibit tendencies towards non-normality. A GOF test for normality is usually conducted on these residuals to determine the adequacy of the prediction model.

The population mean, μ, for this test is 0 and the standard deviation , σ, is assumed to be unknown and has to be estimated from the data. For theoretical consistency, these parameters should be estimated using the maximum likelihood method. Under the null hypothesis, the sample mean, $\bar{X}$, is a maximum likelihood estimator for the population mean. The population standard deviation for a sample of size n can be estimated from

$$\hat{\sigma} = \left[\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n} \right]^{1/2}$$

where X_i represents observation i. This is the maximum likelihood estimator for the population standard deviation given the null hypothesis that the data comes from a normal distribution.

The data is then ordered and grouped into k classes. The determination of class groupings for sample data sets has long been a subject of contention. A relatively robust rule-of-thumb method frequently suggested is for the expected frequency within each class to be at least 5. However, it should be noted that there has been no general agreement regarding this minimum expected frequency. If the expected frequency is too small, adjacent cells can be combined.

For a discrete distribution such as Poisson or binomial, data can be naturally assigned into discrete classes with class boundaries clearly defined based on this rule of thumb. The classes may be set up to ensure that the number of classes should not exceed $n/5$. This is to ensure that the number of observations in each class is not less than 5. The class boundaries can be determined in an equiprobability manner where the probability of a random observation falling into each class is equal and estimated to be $1/k$ for k classes as follows:

$$P\,(x \le x_1) = \frac{1}{k},\ P\,(x \le x_2) = \frac{2}{k}, \ldots,\ P\,(x \le x_{k-1}) = \frac{k-1}{k}.$$

Table 11.2 Class definitions and expected and observed frequencies in each class.

class (i)	Class probability	Class boundary	Expected frequency (E_i)	Observed frequency (O_i)	$(O_i - E_i)^2$
1	0.143	< -0.372	5.71	7	1.653
2	0.143	-0.198	5.71	5	0.510
3	0.143	-0.064	5.71	6	0.082
4	0.143	0.061	5.71	4	2.939
5	0.143	0.195	5.71	6	0.082
6	0.143	0.369	5.71	6	0.082
7	0.143	> -0.369	5.71	6	0.082

As the probability that an observation falls into class i, $p_{i,0}$, can be determined based on the hypothesized distribution with estimated parameters, the class boundaries can be evaluated by inverting this distribution. This procedure can also be used for continuous distributions as illustrated in the example.

The residuals data described in Table 11.1 can be grouped according to classes shown in Table 11.2. The class boundaries are evaluated by inverting the normal cumulative disribution function (CDF) given by

$$\Phi(x) = P(X \leq x) = \int_{-x}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(z-\mu)^2}{2\sigma^2}\right) dz.$$

In Microsoft Excel, this complicated looking function can be inverted using the NORMINV function to derive the class boundaries in Table 11.2. The observed frequencies in each class are also tabulated.

From the data, the X^2 statistic shown in equation (11.1) is evaluated as 0.95. Under the χ^2 distribution with $k - c - 1 = 5$ degrees of freedom the probability of obtaining a value that is at least as large as the computed X^2 is 0.97. This is much larger than 0.05. Hence, the null hypothesis that the data comes from a normal distribution can be retained at the 5% level of significance. The critical chi-square statistic can be computed with the CHIINV function in Microsoft Excel with α and the number of degrees of freedom as the input parameters.

11.4 EMPIRICAL DISTRIBUTION FUNCTION BASED APPROACHES

EDF based approaches essentially rest on the deviations between the empirical distribution function and the hypothesized CDF. Given n ordered observations $x_{(i)}$, an EDF ($F_n(x)$) is defined by

$$F_n(x) = \frac{N\left(x_{(i)} \leq x\right)}{n}, \qquad i = 1, 2, 3, \ldots, n, \tag{11.2}$$

where $N\left(x_{(i)} \leq x\right)$ is the number of ordered observations less than or equal to x. The EDF is also commonly known as the cumulative step function[6] as it can be expressed

as follows:

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \dfrac{r}{n}, & x_{(r)} \le x \le x_{(r+1)}, \\ 1, & x_{(n)} \le x. \end{cases}$$

The development of EDF based approaches initially focused on continuous data. Subsequently, several modifications were developed for discrete and grouped data.

EDF based statistics can be broadly classified into two types, based on how the deviation between the EDF (F_n) and the hypothesized CDF (F_0) is measured.[9] The 'supremum' class of statistics essentially computes the maximum deviations between F_n and F_0. This class includes the well-known Kolmogorov–Smirnov D statistic. The 'quadratic' class computes the following measure of deviation between F_n and F_0:

$$Q^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 \psi(x) d F_0(x), \tag{11.3}$$

where n is the sample size and $\psi(x)$ is a weighting function for the deviations $[F_n(x) - F_0(x)]^2$. This class includes the Cramér--von Mises family of statistics which encompass GOF tests that utilize statistics such as the Cramér–von Mises statistic, the Anderson–Darling statistic and the Watson statistic. Here, the more popular Cramér--von Mises and Anderson–Darling statistics are discussed. Details of the Watson statistic (commonly denoted by U^2), useful for some special cases such as points on a circle, can be found elsewhere.[10] The essential difference between statistics in this class is that the weighting function, $\psi(x)$, is defined differently so as to weight deviations according to the importance attached to various portions of the distribution function.

In this section, the Kolmogorov–Smirnov D statistic is described first followed by the Cramér–von Mises and Anderson–Darling statistics.

11.4.1 Kolmogorov–Smirnov

The Kolmogorov–Smirnov test based on the D statistic is perhaps the best-known EDF based GOF test. The measure of deviation in such a test is essentially the maximum absolute difference between $F_n(x)$ and $F_0(x)$:[6,11,12]

$$D = \sup_x \left[|F_n(x) - F_0(x)| \right].$$

Although this expression for D looks problematic, the asymptotic distribution of this statistic was established by Kolmogorov,[13] and later also by Feller[12] and Doob.[11]

Due to the structure of this statistic, a confidence band can easily be set up such that the true distribution, $F(x)$, lies entirely within the band with a probability of $1 - \alpha$. This can be done as follows. Given any true $F(x)$, if d_α is the critical value of D for test size α,

$$P\{D > d_\alpha\} = \alpha.$$

This probability statement can be inverted to give the confidence statement:

$$P\left\{F_n(x) - d_\alpha \le F(x) \le F_n(x) + d_\alpha, \forall x\right\} = 1 - \alpha. \tag{11.4}$$

Table 11.3 Calculation of absolute deviations.

				$\sup_x[\lvert F_n(x) - F_0(x)\rvert]$	
Residuals	$N(x_{(i)} \le x)$	$F_n(x)$	$F_0(x)$	$\lvert F_n(x_{i-1}) - F_0(x_i)\rvert$	$\lvert F_n(x_i) - F_0(x_i)\rvert$
−0.62	1	0.03	0.039	0.039	0.014
−0.58	2	0.05	0.050	0.025	0.000
−0.57	3	0.08	0.052	0.002	0.023
:	:	:	:	:	:
−0.19	14	0.35	0.294	0.031	0.056
−0.19	15	0.38	0.294	0.056	0.081 D(max)
−0.14	16	0.40	0.345	0.030	0.055
:	:	:	:	:	:
0.55	39	0.98	0.941	0.009	0.034
0.69	40	1.00	0.975	0.000	0.025

Hence, a band can simply be set up of width $\pm d_\alpha$ around sample distribution function $F_n(x)$ such that the true distribution function, $F(x)$, lies entirely within this band. This inversion is allowed because of the measure of deviation and the existence of the distribution for the Kolmogorov–Smirnov D statistic.

The residuals data from helicopter flight times described in Table 11.1 is used to demonstrate the Kolmogorov–Smirnov GOF test. Based on the assumptions in linear regression analysis, the GOF test is for a hypothesized population that is normally distributed with zero mean and unknown standard deviation, σ.

First, the EDF, $F_n(x)$, for each observation is evaluated. The EDF values corresponding to each sample observation are tabulated in Table 11.3. Given that our hypothesized distribution is normal, theoretical values based on this hypothesized normal CDF can be evaluated. These are also shown in Table 11.3. The maximum absolute difference is highlighted in the table, and also shown graphically in Figure 11.1. The band of

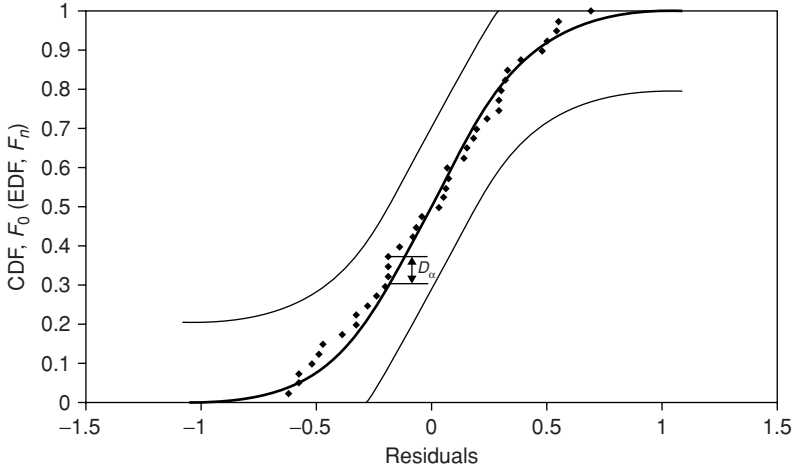


Figure 11.1 Plot of F_n and F_0 against sample observations.

Table 11.4 Critical values of EDF based statistics (Kolmogorov–Smirnov, Anderson–Darling and Cramér–von Mises).

(a) Critical values of D statistic (case where $F_0(x)$ completely specified and mean specified, variance unknown)

Sample size, n	$F_0(x)$ completely specified (from exact distribution)[6]		Mean specified, variance unknown (based on Monte Carlo simulation)[15]	
	$\alpha = 5\,\%$	$\alpha = 10\,\%$	$\alpha = 5\,\%$	$\alpha = 10\,\%$
5	0.565	0.510	–	–
10	0.41	0.368	0.402	0.360
15	0.338	0.304	–	–
20	0.294	0.264	0.288	0.259
25	0.27	0.240	–	–
30	0.24	0.220	–	–
35	0.23	0.210	–	–
50	$1.36/\sqrt{n}$	$1.22/\sqrt{n}$	0.185	0.165
100			0.132	0.118
>100			$1.333/\sqrt{n}$	$1.190/\sqrt{n}$

(b) Critical values for A^2 and W^2 statistics for Anderson–Darling and Cramér–von Mises GOF test ($n \geq 5$)[15]

Case when $F_0(x)$ completely specified

	Modifications	Critical values Modified W^* and A^2 statistic	
		$\alpha = 5\,\%$	$\alpha = 10\,\%$
W^2	$W^* = (W^2 - 0.4/n + 0.6/n^2)(1.0 + 1.0/n)$	0.461	0.347
A^2	–	2.492	1.933

Case when mean specified, variance unknown and estimated from $s^2 = \sum_i (x_i - \mu)^2 / n$

	Critical values at α level of significance	
	$\alpha = 5\,\%$	$\alpha = 10\,\%$
W^2	0.443	0.329
A^2	2.323	1.76

absolute differences where the null hypothesis can be retained at the 5% level of significance is shown in Figure 11.1 as well. This 'acceptance' band is computed from the critical values for the absolute difference based on equation (11.4). The critical values d_α are given by Barnes and Murdoch.[14] The critical values for the case where $F_0(x)$ is completely specified and the case where the mean is specified and the variance unknown given in Table 11.4(a).[6,15]

For a sample size of $n = 40$, the critical value for the maximum absolute difference, D, at the 5 % level of significance for the case where the population mean is known but the standard deviation is unknown is approximately 0.206. Since the maximum absolute difference for this sample is 0.081, we do not have sufficient evidence to reject the null hypothesis that this data set comes from a normal distribution at the 5% level of significance. It can also be observed from Figure 11.1 that the sample observations all fall within the acceptance band.

In some GOF tests, neither the population mean, μ, nor the standard deviation, σ, is specified and they must be estimated from the sample observations. In this case, a modified Kolmogorov–Smirnov statistic, D^*, can be used. The modifications to the original D statistic and corresponding critical values for this case are given by Stephens.[15] These modifications are derived from points for finite n obtained through Monte Carlo simulations instead of theoretical results.

11.4.2 Cramér–von Mises

The Cramér–von Mises test statistic for GOF testing belongs to the class of quadratic measure based EDF statistics, based on the measure of deviation shown in equation (11.3). In this case the weighting function, $\psi(x)$, is assumed unity. Hence, from (11.3), the Cramér–von Mises test statistic is given by:

$$W^2 = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x).$$

In practical implementations, the following formula based on ordered observations $(x_{(i)})$ can be used instead:

$$W^2 = \sum_{i=1}^{n} \left[F_0(x_{(i)}) - \frac{i - 0.5}{n} \right]^2 + \frac{1}{12n}. \tag{11.5}$$

The residuals data from helicopter flight times shown in Table 11.1 is used to illustrate the procedure for conducting a the Cramér–von Mises GOF test. This test is again for a null hypothesis that the data comes from a normal distribution with population mean known but standard deviation unknown.

The W^2 statistic computed from equation (11.5) is less than the theoretical critical values of the W^2 statistic at the 5% significance level (the table of critical values for the W^2 statistics is given in Table 11.4(b)).[15] Hence, the specified hypothesis that the sample data comes from a normal distribution cannot be rejected at the 5% level of significance. Unlike the GOF test for residuals in this example where the population mean is known, the population parameters are often unknown in practice and have to be estimated from the sample data. For these cases, modifications have been proposed to the original Cramér–von Mises statistic.[9,15] These modifications are reproduced in Table 11.5.

Table 11.5 Modifications of A^2 and W^2 statistics for mean, μ, and variance, σ^2, unknown* and the corresponding critical values [9].

EDF based Statistic	Scaling	Significance level (α) 0.05	0.10
A^2	$A^2(1.0 + 0.75/n + 2.25/n^2)$	0.752	0.631
W^2	$W^2\ (1.0 + 0.5/n)$	0.126	0.104

*Mean, μ, estimated from $\bar{x}$, and standard deviation, σ, estimated from

$$s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$$

11.4.3 Anderson–Darling

Another popular EDF based GOF test for normality is based on the Anderson–Darling test statistic. Like the Cramér-von Mises test statistic, it is based on the quadratic measure of deviation shown in (11.3), but with weighting function

$$\psi(x) = \left\{F_0(x)[1 - F_0(x)]\right\}^{-1}.$$

Such a weighting function results in deviations in the tails of distributions (when $F_0(x) = 0$ or 1) being weighted more heavily. Hence, this test is more sensitive to deviations in the tails of distributions. For ordered observations ($x_{(i)}$) of n observations, the Anderson–Darling test statistic, A^2, is given by

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{\ln F_0(x_{(i)}) + \ln[1 - F_0(x_{(n+1-i)})]\right\}. \tag{11.6}$$

The limiting distribution of A^2 is given in Anderson and Darling's original paper.[16] As this statistic requires the use of specific distributions in calculating the critical values, it possesses the advantage of being a more sensitive test. However, the critical values have to be calculated for each distribution. This test can be used for normal, lognormal, exponential, Weibull, extreme value type I and logistic distributions.

We return to the residuals data set in Table 11.1 to demonstrate the Anderson–Darling GOF test. As before, the hypothesis that this data set comes from a normal distribution is tested with the assumption that the population mean is zero and the standard deviation unknown and estimated from data. Using equation (6), A^2 is found to equal 0.262. This is lower than the critical A^2 value at the 5% significance level of 2.323 shown in Table 11.4(b).[15] For completeness, modifications to the A^2 statistic in order to deal with GOF tests where the population mean and standard deviation are estimated from data are given in Table 11.5.[9] Critical values for the modified A^2 statistics for this case are also reproduced in Table 11.5 together with the scaling for a modified Cramér–von Mises W^2 statistic and its corresponding critical values for cases where the population mean and standard deviation are unknown.

11.5 REGRESSION-BASED APPROACHES

11.5.1 Probability plotting

There are many types of graphical procedures available for assessing goodness of fit. The simplest of these involve plotting the EDF together with the hypothesized CDF against the corresponding ordered sample data, $x_{(i)}$. An example is shown in Figure 11.2(a). Such a plot can be used to visually examine distributional characteristics (such as skewness and kurtosis) and to detect outliers and the presence of contamination (mixtures of distributions).[9] However, the difficulty of assessing the GOF given the curved nature of distribution functions has been widely acknowledged.

Probability plotting is an elegant graphical procedure which offers the advantage of judging the goodness-of-fit based on a straight line. This is made possible by transforming the EDF axis using special probability plotting paper such that the sample observations should fall roughly in straight line if the hypothesized distribution adequately represents the data. Different hypothesized models demand different types of probability plotting papers. If the points on such a plot fall along a straight line on the probability plotting paper, the underlying probability distribution can be considered adequate for modeling the data. An example of probability plotting is shown in Figure 11.2(b) alongside the simple procedure described in the previous paragraph. It is relatively easier to judge the quality of a linear fit compared to a higher-order nonlinear fit.

Probability plotting has been widely acknowledged as an extremely versatile and useful tool as it not only provides a pictorial representation of data, but also allows nonparametric estimation of the percentiles and other model parameters of the underlying distributions. Furthermore, it can readily be used for censored observations as commonly encountered in life-data analysis for reliability and survival analysis applications.

In recent years, many statistical software programs have been developed to automate the probability plotting process. These include MINITAB, Statgraphics, and JMP. In the absence of such automated means and probability plotting papers, normal graph papers can be used, given the existence of simple linearizing transformations. Such transformations are discussed here and demonstrated through the use of the

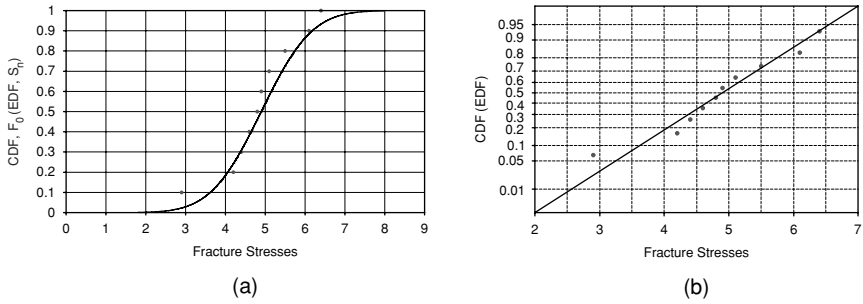


Figure 11.2 (a) Simple graphical goodness-of-fit comparison. (b) Graphical goodness-of-fit comparison using probability plotting on normal probability paper.

commonly available Microsoft Excel platform. The fundamental concepts underlying this transformation are discussed first, followed by a worked example using Minitab and Microsoft Excel to demonstrate the usefulness of these techniques.

11.5.1.1 Fundamental concepts in probability plotting

Probability plots are essentially plots of the quantile values against the corresponding ranked observations ($x_{(i)}$). Hence, in general, they can be represented in the following functional form:

$$x_{(i)} = F^{-1}[p(x_{(i)}; n)].$$

Here, the probability value, $p(x_{(i)}; n)$, for evaluating the quantile values can be estimated from

$$p(x_{(i)}; n) \approx \frac{i - 0.5}{n}. \tag{11.7}$$

The underlying principles of probability plotting are based on the expected value of an ordered observation, $E(x_{(i)}; n)$.[7] Each ordered random observation in a sample of size n corresponds to one such expected value. In a single sample of size n, each ordered random observation $x_{(i)}$ is a one-sample estimate of $E(x_{(i)}; n)$. Hence, when each of these ordered random observations are plotted against its expected value, they should approximately lie along a straight line through the origin with slope 1.

The expected value of an ordered observation in a sample is distribution-dependent. For most distributions, the expected value of the ith observation can be estimated from

$$E(x_{(i)}; n) \approx F^{-1}\left\{\frac{i - c}{n - 2c + 1}\right\}, \qquad c \in [0, 0.5].$$

This is essentially the $((i - c)/(n - 2c + 1))$th quantile of the distribution evaluated at $x_{(i)}$. The constant c is a function of both the hypothesized distribution and the sample size. A value of $c = 0.5$ is generally acceptable for a wide variety of distributions and sample sizes, giving the estimation shown in equation (11.7). For the uniform distribution, c is taken to be zero and the expected value is given by

$$E(x_{(i)}; n) \approx F^{-1}\left(\frac{i}{n + 1}\right).$$

11.5.1.2 Linearizing the CDF

In the absence of convenient probability plotting papers or statistical software offering facilities for automated plotting, normal graph papers can be used in conjuction with linearization of the hypothesized CDF. The CDFs of many common distributions can be linearized by taking advantage of the structure of the quantile function. Such linearization transforms the data and allows it to be plotted against the cumulative percentage of observations or CDF (F_0) on normal graph papers. If the corresponding plotted points fall roughly on a straight line, similar assessments that the data can be adequately described by the normal distribution can be made.

The CDF of a normally distributed random variable $x(x \sim N(\mu, \sigma))$, in terms of the standard normal CDF, $\phi(\cdot)$, is given by

$$F(x) = \phi\left(\frac{x-\mu}{\sigma}\right).$$

Hence, the quantile function, $x_{(i)}$, for the $[F(x_{(i)})]$th quantile corresponding to the ith ordered observation $(x_{(i)})$ in terms of the standard normal CDF, $\Phi(\cdot)$, is given by

$$x_{(i)} = \mu + \Phi^{-1}[F(x_i)]\sigma.$$

In addition, consequent to the rank-ordering of sample observations, the expected value of ordered observations at each $[F(x_{(i)})]$th quantile can be approximated by

$$E(x_{(i)}; n) \approx \Phi^{-1}\left[\frac{i-0.5}{n}\right].$$

Each ordered observation can then be plotted against the expected value or $\Phi^{-1}(\cdot)$. This should give a straight line with intercept μ and slope σ if the normal distribution appears adequate for representing the data.

Here, MINITAB and Microsoft Excel are used to generate the normal probability plots. However, such plots can be easily generated with normal probability plotting papers or standard graph papers as already discussed.

11.5.1.3 Using MINITAB and Excel

To conclude this section, MINITAB and Excel were used to produce probability plots for data set in Table 11.1. Figure 11.3 shows the probability plot generated by MINITAB together with the Anderson–Darling statistic generated for this set of residuals; also plotted are the corresponding confidence limits. As none of the residuals are outside the 95 % confidence interval bands, the null hypothesis that the residuals come from a normal distribution can thus be retained at the 5 % level of significance. In fact, based on the A^2 statistic, the level of significance can be as high as 25 % as shown by the p-value in Figure 11.3.

In Excel, the inverse of the standard normal CDF, $\Phi^{-1}(\cdot)$, can be easily evaluated using the NORMSINV function. A probability plot can then be generated as shown in Figure 11.4. A best-fit trendline can be drawn through the data points using Excel graphing tools. The intercept of this linear fit provides an estimate of the mean and its slope provides an estimate of the standrad deviation.

11.5.2 Shapiro–Wilk Test

Given the above discussion on the importance of linearity in probability plots, statistics have been developed to measure such linearity. The Shapiro–Wilk statistic is one of these. It is commonly used in GOF tests for normality and lognormality as its underlying distribution based on the normal distribution.

The Shapiro–Wilk GOF test is relatively powerful GOF test for normality and is usually recommended for cases with limited sample data. The following test statistic is computed:

$$W = \frac{b^2}{S^2}. \tag{11.8}$$

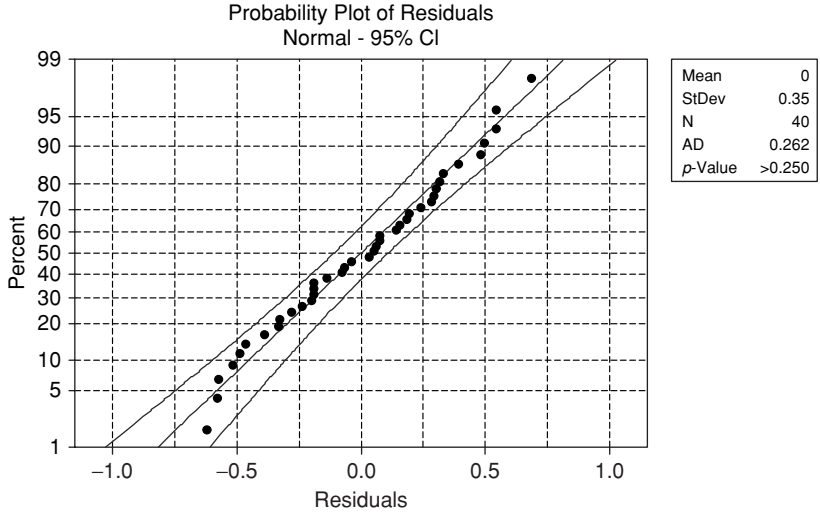


Figure 11.3 Probability plot of residuals using MINITAB.

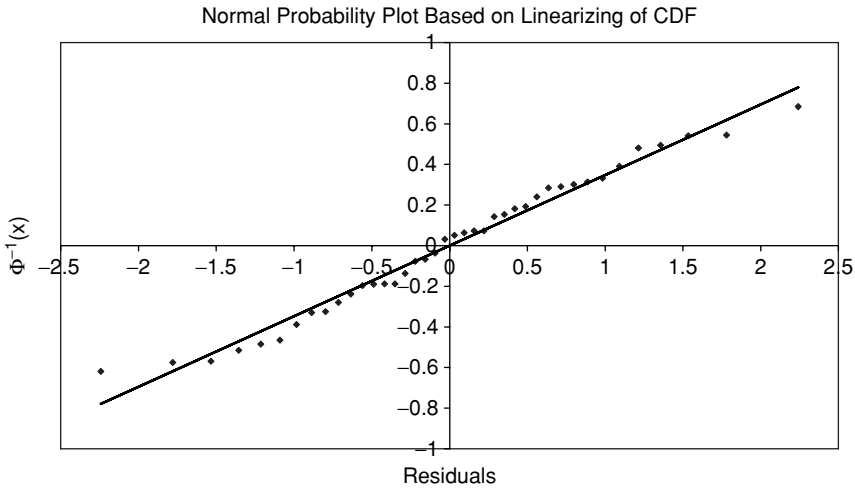


Figure 11.4 Probability plot of residuals using Excel.

where

$$S^2 = \sum_{i=1}^{n} (x_{(i)} - \bar{x})^2,$$

$$b = \sum_{i=1}^{k} a_{i,n}(x_{(n-i+1)} - x_{(i)}),$$

in which the $x_{(i)}$ are ordered observations of rank i; $k = n/2$ when n is even $(n-1)/2$ when n is odd; the $a_{i,n}$ are constants generated from the moments of the standard

normal order statistics of size n (see Table 11.6(a)).[17,18] Small values of W are evidence of departure from normality.

Using the data in Table 11.1, a W statistic of 0.966 is obtained. This is greater than the critical value of the W statistic at the 5 % significance level for $n = 40$ (see Table 11.6(b)).[17,18] Hence we have insufficient evidence to reject the null hypothesis that the data did not originate from a normal distribution. For this computation, the percentage points of the W statistics and constants, $a_{i,n}$, can be obtained from Table 11.6(a).

11.6 FISHER'S CUMULANT TESTS

Fisher's Cumulant tests are GOF tests based on the standardized third and fourth moments of the distribution. Such tests are founded on the recognition that deviations from normality can be characterized by the standardized third and fourth moments of the distribution, known as the skewness and kurtosis, respectively.

Roughly speaking, skewness measure the symmetry of the distribution about the mean and kurtosis measures the relative peakedness or flatness of the distribution. The normal distribution is symmetrical about the mean, and median and mode coincide with the mean; it is said to have zero skewness. Distributions with an asymmetric tail extending towards higher values are said to possess *positive skewness* (or *right-handed skew*) and distributions with an asymmetric tail extending towards lower values are said to possess *negative skewness* (or *left-handed skew*).

Kurtosis characterizes the relative peakedness of a distribution compared to a normal distribution. A normal distribution having a kurtosis of 0 is described as *mesokurtic*. A *leptokurtic* distribution with positive kurtosis has a relatively more peaked shape than the normal distribution. A *platykurtic* distribution with negative kurtosis has a relatively flatter shape compared to the normal distribution. Kurtosis is relevant only only for symmetrical distributions.

Apart from a characterization of skewness and kurtosis using the standardized moments, these distribution characteristics can also be described in terms of the cumulants of the distribution. The test statistics for the corresponding GOF tests have been described in terms of the following first four sample cumulants which are also called the Fisher's K-statistics. Here Fisher's cumulant GOF tests are described directly in terms of the sample skewness and sample kurtosis whose distribution under normality assumption can be found in the book by Pearson and Hartley.[18]

In Fisher's cumulant GOF test for normality, the test statistic for assessing the skewness of the distribution is defined as the *sample skewness coefficient*,

$$\hat{\gamma}_1 = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - x)^3}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}.$$

A simple GOF test of normality would reject large values of $|\hat{\gamma}_1|$. However, since this test is targeted specifically at the skewness of distributions, it will have poor power against any alternative distributions with $\gamma_1 = 0$ (zero skewness) which are nonnormal.

Table 11.6 Coefficients and critical values for the Shapiro--Wilks GOF test.[17,18]

(a) Coefficients $a_{i,n}$ computing the W statistics

	Sample size, n					
i	5	10	20	30	40	50
1	0.6646	0.5739	0.4734	0.4254	0.3964	0.3751
2	0.2413	0.3291	0.3211	0.2944	0.2737	0.2574
3	0.0000	0.2141	0.2565	0.2487	0.2368	0.2260
4	—	0.1224	0.2085	0.2148	0.2098	0.2032
5	—	0.0399	0.1686	0.1870	0.1878	0.1847
6	—	—	0.1334	0.1630	0.1691	0.1691
7	—	—	0.1013	0.1415	0.1526	0.1554
8	—	—	0.0711	0.1219	0.1376	0.1430
9	—	—	0.0422	0.1036	0.1237	0.1317
10	—	—	0.0140	0.0862	0.1108	0.1212
11	—	—	—	0.0697	0.0986	0.1113
12	—	—	—	0.0537	0.0870	0.1020
13	—	—	—	0.0381	0.0759	0.0932
14	—	—	—	0.0227	0.0651	0.0846
15	—	—	—	0.0076	0.0546	0.0764
16	—	—	—	—	0.0444	0.0685
17	—	—	—	—	0.0343	0.0608
18	—	—	—	—	0.0244	0.0532
19	—	—	—	—	0.0146	0.0459
20	—	—	—	—	0.0049	0.0386
21	—	—	—	—	—	0.0314
22	—	—	—	—	—	0.0244
23	—	—	—	—	—	0.0174
24	—	—	—	—	—	0.0104
25	—	—	—	—	—	0.0035

(b) Critical values for W statistics

	Critical values at α-level of significance	
Sample size, n	$\alpha = 5\,\%$	$\alpha = 5\,\%$
5	0.762	0.806
10	0.842	0.869
20	0.905	0.92
30	0.927	0.939
40	0.94	0.949
50	0.947	0.955

Table 11.7 Critical values of moment statistics for Fisher's cumulant GOF test.[18]

	Sample skewness (γ_1)		Sample skewness (γ_2)	
Sample size, n	$\alpha = 5\,\%$ Upper*	$\alpha = 10\,\%$ Upper*	$\alpha = 5\,\%$ Upper (Lower)	$\alpha = 10\,\%$ Upper (Lower)
20	0.940	0.772	1.68 (−1.27)	1.18 (−1.17)
25	0.866	0.711	–	–
30	0.806	0.662	1.57 (−1.11)	1.12 (−1.02)
35	0.756	0.621	–	–
40	0.714	0.588	1.46 (−1.01)	1.06 (0.93)
45	0.679	0.559	–	–
50	0.647	0.534	1.36 (−0.94)	1.00 (−0.85)
100	0.470	0.390	1.03 (−0.73)	0.77 (−0.65)

*Lower limits here are equivalent to the corresponding upper limits with negative sign.

The test statistic for assessing the kurtosis of the distribution is defined by the *sample kurtosis*,

$$\hat{\gamma}_2 = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} - 3.$$

Percentage points for different critical values of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ for these GOF tests can be found in Table 11.7.[18]

For large samples ($n > 100$), the following test statistic can be used instead of the sample skewness:

$$Z_1 = \frac{\hat{\gamma}_1}{\sqrt{6/n}}.$$

Similarly, for large sample size, the following test statistic can be used in place of the sample kurtosis:

$$Z_2 = \frac{\hat{\gamma}_2 - 3}{\sqrt{24/n}}.$$

Both of these statistics can be approximated by a standard normal distribution. Consequently, there is a combined test which simultaneously takes into account both the skewness and kurtosis of the distribution. This is given by

$$X^2 = Z_1^2 + Z_2^2,$$

which is approximately χ^2 distributed with 2 degrees of freedom.

The skewness and kurtosis measures computed for the data in Table 11.1 are −0.020 and −0.933, respectively. The critical values at 5% significance level for a sample size of $n = 40$ can be obtained from Table 11.7. Two-sided $\hat{\gamma}_1$ and $\hat{\gamma}_2$ tests both yield p-values that are greater than 0.05. Hence, the null hypothesis that the data comes from a normal distribution cannot be rejected at the 5 % level of significance.

The combined X^2 statistics is 1.452. This is lower than the critical χ^2 value at the 5 % significance level, hence the null hypothesis cannot be rejected at the 5 % significance

level. Note, however, that this combined test is not recommended for samples of size 40 due to its poor power performance; it is given here for purposes of illustration only.

11.7 CONCLUSION

In this chapter, the generic concepts behind GOF tests together with several important and popular GOF tests have been discussed. Such tests are essential for the appropriate statistical modeling of real-world data. This is a critical link in the Six Sigma approach prior to any process of product optimization exercise which relies on rigorous statistical models.

REFERENCES

1. Gao, Y., Lam, S.W. and Tang, L.C. (2002) A SWOT analysis to Six Sigma Strategy. *Proceedings of the 8th Asia Pacific Quality Organization Conference*, pp. 197–207.
2. Goh, T.N., Tang, L.C., Lam, S.W. and Gao, Y. (2005) Core Six Sigma: a self assessment. *4th Sino-Korea Bilateral Symposium on Quality*, pp. 3–8.
3. Neyman, J. and Pearson, E.S. (1933) On the problem of the most efficient tests of statistical hypothesis. *Philosophical Transactions of the Royal Society, Series A*, 231, 289–337.
4. Cochran, W. (1952) The chi square test of goodness of fit. *Annals of Mathematical Statistics*, 23, 315–345.
5. Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have risen from random sampling. *Philosophical Magazine, Series 5*, 50, 157–172.
6. Massey, F.J. (1951) The Kolmogorov–Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46, 68–78.
7. Hahn, G.J. and Shapiro, S.S. (1967) *Statistical Models in Engineering*. New York: John Wiley & Sons, Inc.
8. Box, G.E.P. and Liu, P.Y.T. (1999) Statistics as a catalyst to learning by scientific methods, Part 1. An example. *Journal of Quality Technology*, 31, 1–15.
9. D'Agostino, R.B. and Stephens, M. (1986) *Goodness-of-Fit Testing*. New York: Marcel Dekker.
10. Watson, G.S. (1961) Goodness-of-fit tests on a circle. Biometrika, 48, 109–114.
11. Doob, J.L. (1949) Heuristic approach to the Kolmogorov–Smirnov theorems. *Annals of Mathematical Statistics*, 20, 393–403.
12. Feller, W. (1948) On the Kolmogorov–Smirnov limit theorems for empirical distributions. *Annals of Mathematical Statistics*, 19, 177–189.
13. Kolmogorov, A. (1933) Sulla determinazione empirica di una legge disitribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, 1–11.
14. Murdoch, J. and Barnes, J.A. (1998) *Statistical Tables*. New York: Palgrave.
15. Stephens, M.A. (1974) EDF statistics for goodness of fit and some comparisons, *Journal of the American Statistical Association*, 69, 347, 730–737.
16. Anderson, T.W. and Darling, D.A. (1952) Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *Annals of Mathematical Statistics,* 23, 193–212.
17. Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). Biometrika, 52, 591–611.
18. Pearson, E.S. and Hartley, H.O. (1976) *Biometrika Tables for Statisticians*, Vol. 1. London: Biometrika Trust.

12

Introduction to the Analysis of Categorical Data

L. C. Tang and S. W. Lam

Effective statistical analysis hinges upon the use of appropriate techniques for different types of data. As Six Sigma evolves from its original applications in manufacturing environments to new-found applications in transactional operations, categorical responses increasingly become the norm rather than the exception. In this chapter two basic schemes, contingency tables and logistic regression, for the analysis of categorical data are presented. These techniques can easily be implemented in Excel or statistical software such as MINITAB. Two case studies, one for each of these methods, are also given to illustrate their application.

12.1 INTRODUCTION

An important statistical procedure in Six Sigma implementation is the statistical modeling of relationships between the key process input variables (KPIVs) and key process output variables (KPOVs). Linear regression techniques based on ordinary least squares (OLS) have always been the tool of choice in most Six Sigma applications for manufacturing processes. Generally, when the KPOVs are continuous random variables, the usual OLS regression assumptions are embedded in the analysis. However, when the KPOVs are not continuous but measured on scales comprising distinct categories, these OLS assumptions are violated. Such situations are relatively more pervasive in the modeling of relationships between KPIVs and KPOVs for transactional processes. Fortunately, a broad family of statistical tools and techniques has been specially developed to deal with such cases. These techniques are applicable regardless of whether the associated KPIVs are measured on discrete categorical or

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

continuous scales. This chapter attempts to provide a practical introduction to such techniques.

In statistical terminology, KPOVs and KPIVs are commonly known as *response* and *explanatory* variables, respectively. The use of 'KPOV' and 'KPIV' in typical Six Sigma terminology tends to imply a causal relationship between the input and output variables, whereas the terms 'response' and 'explanatory' have more generic connotations. In many situations, the main purpose of studies conducted for transactional or manufacturing processes with qualitative responses may be to analyze the generic associations between qualitative response and explanatory variables. The ability to model causal relationships typically depends on the sampling design, a matter which is beyond the scope of this chapter. In any case, models that can describe the causal relationships between KPIVs and KPOVs are specialized cases of the statistical models described in this chapter. In view of this, the present chapter refers to response and explanatory variables instead of KPOVs and KPIVs, respectively.

Categorical response variables are frequently encountered in many real-world processes and are particularly well suited to data from transactional processes. Qualitative responses such as the degree of customer satisfaction with a particular transactional process are usually measured on a categorical scale such as 'dissatisfied', 'neutral', and 'satisfied'. Such data may arise from customer satisfaction surveys, and the spectrum of response can be widened or narrowed depending on the survey design. In certain situations categorical data may also arise more artificially due to the high cost of obtaining continuous data. It should be noted that resorting to regression techniques based on OLS assumptions for analyzing categorical response data may result in highly flawed conclusions. Typically, categorical response data contains less information and is therefore, more difficult to analyze, and the statistical quality of results is usually not as good as those obtained from continuous data.

In certain situations, the categorical response or explanatory variables possess natural ordering information. Such ordering information can be leveraged on to provide more precise and sensitive statistical inferences. Data with such ordinal information is a special case of categorical data and is typically known as *ordinal* data. The customer satisfaction scale described in the previous paragraph is one example. Other examples include the response to a particular medical treatment in a clinical trial and the response to certain doping processes in metal treatment procedures, where the response can be classified into ordered categorical scales such as 'not effective', 'effective', and 'very effective'. In contrast to ordinal categorical variables, categorical variables with no inherent ordinal information are also pervasive. Such variables are known as *nominal* variables; examples include race, which can be classified as to 'Chinese', 'Malay', 'Indian', 'Eurasian', etc., and color, consisting of 'red', 'green', 'blue', etc. Analyses of nominal data are more generic than those of ordinal categorical data. Both types of categorical data analysis procedure are introduced in this chapter.

There are generally two broad streams of development in the analysis of categorical data: procedures based on the use of contingency tables and those based on generalized linear modeling. Techniques derived from the use of contingency tables were the original statistical tools used in categorical data analysis; these are discussed in the next section. This is followed by a case study demonstrating the use of these fundamental techniques. Later progress in linear modeling techniques led to the development of generalized linear modeling techniques for categorical data analysis;

one member of this family of techniques that is particularly prevalent and effective is logistic regression. This is discussed for cases of single and multiple categorical or continuous explanatory variables in Section 12.4.

12.2 CONTINGENCY TABLE APPROACH

Categorical data can typically be presented in a tabular format when both the response and explanatory variables are categorical in nature, or can be defined in distinct categories. Variables which are categorical in nature are commonly referred to as *factors*, and the different categories commonly referred to as *factor levels*. In many situations, the data in contingency tables are the frequency counts of observations occurring for each possible factor-level combination.

Table 12.1 shows a typical two-way contingency table for a simple situation with only two categorical variables, X and Y, with I and J levels, respectively. Each variable n_{ij} ($i = 1, \ldots, I$; $j = 1, \ldots,$ J) in the table shows the frequency of counts in each (i, j) factor-level combination. For each row (column) the marginal sums are shown in the 'Total' row (column). The total sample size is denoted by n.

A number of statistical measures and procedures have been proposed to assess the association or relationship between variables in categorical data analysis. Statistical measures such as sample proportions, relative risks and odds ratios can be used in the case of binary variables in two-way contingency tables (see the case study in Section 12.3).

Another key method is to use a rigorous statistical hypothesis test. Let π_{ij} denote the probability of an observation belonging to category $X = i$ and $Y = j$. The π_{ij} thus define the *joint probability distribution* of X and Y. Denote by π_{i+} the marginal probability of an observation belonging to category $X = i$ and by π_{+j} the marginal probability of an observation belonging to category $Y = j$. A typical hypothesis test for a two-way contingency table with only one response and one explanatory variable is as follows:

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j} \quad \text{vs.} \quad H_1: \pi_{ij} \neq \pi_{i+}\pi_{+j}, \qquad \text{for all } i \text{ and } j.$$

The null hypothesis H_0 states that the variables X and Y are statistically independent. When this holds the probability of an observation falling in any particular column is independent of which row that observation belongs to. This results in the

Table 12.1 Two-way contingency table.

		Y				
		Level 1	Level 2	...	Level J	Total
X	Level 1	n_{11}	n_{12}	...	n_{1J}	n_{1+}
	Level 2	n_{21}	n_{22}	...	n_{2J}	n_{2+}
	:	:	:		:	:
	Level I	n_{I1}	n_{I2}	...	n_{IJ}	n_{I+}
	Total	n_{+1}	n_{+2}		n_{+J}	n

probabilistic statement that the joint probability is equivalent to the product of the marginal probabilities. The conclusion of this hypothesis test applies in more general two-variable cases without reference to any distinction between response and explanatory variables. Specifically, if the null hypothesis holds, the probability of an observation falling in any row is independent of which column that observation already belongs to. Hence, if the null hypothesis is rejected, there is evidence suggesting the presence of a relationship between the variables.

The marginal probabilities, π_{i+} and π_{+j}, can be estimated by the sample marginal probabilities, p_{i+} and p_{+j}, respectively:

$$p_{i+} = \frac{n_{i+}}{n}, \tag{12.1}$$

$$p_{+j} = \frac{n_{+j}}{n}. \tag{12.2}$$

Under the null hypothesis, the joint probability in each cell, π_{ij}^0, is given by the product of these marginal probabilities and can be estimated by the sample estimators for these marginal probabilities as follows:

$$p_{ij}^0 = p_{i+}p_{+j} = \frac{n_{i+}}{n}\frac{n_{+j}}{n}. \tag{12.3}$$

With the estimated joint probabilities under the null hypothesis, comparisons can be made with the estimated joint probabilities of each cell obtained from the actual data. Statistical tests based on the expected frequencies can be used to implement these comparisons. The estimated expected frequency for each (i, j)th cell under the null hypothesis is given by

$$\hat{\mu}_{ij} = np_{i+}p_{+j} = \frac{n_{i+} \cdot n_{+j}}{n}. \tag{12.4}$$

Given that the actual cell frequencies are n_{ij}, the following Pearson X^2 statistic can be used to assess the independence between the two variables:

$$X^2 = \sum_{i,j} \frac{n_{ij} - \hat{\mu}_{ij}}{\hat{\mu}_{ij}}. \tag{12.5}$$

Another statistic based on likelihood ratios can also be used:

$$G^2 = 2\sum_{i,j} n_{ij} \ln \frac{n_{ij}}{\hat{\mu}_{ij}}. \tag{12.6}$$

For both of these statistics, the large-sample reference distribution is χ^2 with $(I-1) \times (J-1)$ degrees of freedom. The null hypothesis is rejected when the computed statistic exceed the critical value at a given significance level.

More detailed information can be obtained by looking at the contribution to the overall X^2 and G^2 statistics from each variable combination (or cell) and the difference between the observed and expected cell frequencies in each cell. The differences between observed and expected frequencies in each cell are also known as the *residuals*.

In order to compensate for scaling effects, a more appropriate measure for comparison of residuals is given by the adjusted residuals,

$$R_{\text{adj}} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1 - p_{i+})(1 - p_{+j})}}. \tag{12.7}$$

For the test of independence between two variables, X^2 and G^2 statistics are sufficient for nominal data. However, if ordinal information is available, analysis based on the X^2 and G^2 statistics may not be as sensitive as a test which takes into account such information. This ordinal information can be derived from a natural ordering of the levels of the variables. When the association has a positive or negative trend over the range of ordered categories, tests which leverage on the obvious presence of such ordinal information are more sensitive to departures from the null hypothesis. A statistic that encapsulates the ordinal information is given by

$$M^2 = (n - 1)r^2, \tag{12.8}$$

where r is the Pearson product-moment correlation between X and Y, and n is the total sample size. The null hypothesis using this statistic is that of independence between variables X and Y, and the alternative hypothesis states the presence of significant correlations between these two variables. In this test, the M^2 statistic follows a null χ^2 distribution with 1 degree of freedom. r, which accounts for the ordinal information underlying the categories, can be calculated as follows:

$$r = \frac{\sum_{i,j} u_i v_j n_{ij} - \left(\sum_i u_i n_{i+}\right)\left(\sum_j v_j n_{+j}\right)\Big/ n}{\sqrt{\left[\sum_i u_i^2 n_{i+} - \frac{\left(\sum_i u_i n_{i+}\right)^2}{n}\right]\left[\sum_j v_j^2 n_{+j} - \frac{\left(\sum_j v_j n_{+j}\right)^2}{n}\right]}}, \tag{12.9}$$

where the u_i are the scores of the ith rows, with $u_1 \leq u_2 \leq u_3 \leq \ldots \leq u_I$, and the v_i are the scores of the jth columns, $v_1 \leq v_2 \leq v_3 \leq \ldots \leq v_J$.

From equation (12.9), the frequency counts can be observed to be weighted by the scores of the respective rows and columns. For most data sets, the choice of scores has little effect on the result if they are reasonably well chosen and equally spaced. However, in some cases, the imbalance in frequency counts over the categories may give different results for different scores. In such cases, sensitivity analysis can be conducted to assess these differences for different scoring system. Other approaches suggested in literature include the use of data to assign scores automatically.[1] However, such automatic scoring systems may not be appropriate for all circumstances. It is usually better to leverage on reasonable domain knowledge in selecting scores that reflect the differences between categories.

A two-way $I \times J$ contingency table can be generalized to a three-way $I \times J \times K$ contingency table and even multi-way contingency tables. An example of an $I \times J \times K$ three-way contingency table is shown in Table 12.2. In this chapter, only two-way contingency tables for both nominal and ordinal categorical data are dealt with. For multi-way contingency tables involving more variables, the reader is advised to refer to Agresti.[1]

Table 12.2 Three-way contingency table.

		Y			
Z	X	Level 1	Level 2	...	Level J
Levels	*Levels*				
1	1	n_{111}	n_{112}	...	n_{11J}
	2	n_{121}	n_{122}	...	n_{12J}
	⋮	⋮	⋮		⋮
	I_1	n_{1I1}	n_{1I2}	...	n_{1IJ}
⋮	1	n_{i11}	n_{i12}	...	n_{i1J}
	2	n_{i21}	n_{i22}	...	n_{i2J}
	⋮	⋮	⋮		⋮
	I_i	n_{iI1}	n_{iI2}	...	n_{iIJ}
K	1	n_{K11}	n_{K12}	...	n_{K1J}
	2	n_{K21}	n_{K22}	...	n_{K2J}
	⋮	⋮	⋮		⋮
	I_K	n_{KI1}	n_{KI2}	...	n_{KIJ}

12.3 CASE STUDY

In this section a case study is described to demonstrate the use of contingency table techniques for categorical data analysis. The case study is based on data obtained from an annual national survey of entities engaged in research and development. The actual survey was much broader in scope and sought to obtain a variety of information related to the R&D activities of private and public organizations. Data related to the sectoral employment of research scientists/engineers (RSEs) based on their qualification level is extracted in order to assess whether having a PhD is associated with a better chance of employment in the private sector. The actual survey is not specifically targeted to achieve this objective. However, the data set is simple enough for demonstrating the contingency table approach here.

Table 12.3 cross-classifies 18 935 survey respondents by qualifications and employment sector. Here, the response variable is the number of RSEs employed in the private sector and the explanatory variable is the qualification level (with or without PhD). The association we are concerned here with is the conditional distribution of employment in the private sector, given the qualifications of these RSEs. This is the simplest case in the contingency table type of categorical data analysis where there are only two categories for each of the two variables, resulting in a 2×2 contingency table. Response variables with only two categories are also called *binary* response variables.

The X^2 and G^2 statistics can be calculated using equations (12.5) and (12.6) as 3847 and 3876. The adjusted residuals for each cell can be calculated using (12.7). The X^2 and G^2 statistics and their corresponding probability values describe the evidence against the null hypothesis. In this particular example, the overall computed X^2 and

Table 12.3 Contingency table for the study on the sectoral employment of RSEs.

	Employment in		
RSEs	Private sector	Public sector	Total
With PhD	781	3 282	4 063
Without PhD	10 815	4 057	14 872

G^2 statistics are very much larger than the critical χ^2 statistic at the 95% level of significance. Hence, there is very strong evidence to suggest that there is a relationship between the PhD qualifications and sectoral employment status of RSEs. The full computation of X^2 contributions and adjusted residuals for each cell is shown in Table 12.4. From the adjusted residuals there appear to be significantly more RSEs without PhDs in the private sector than predicted by a hypothesis of independence. Conversely, there appear to be significantly more PhD-level RSEs in the public sector than would occur if the sectoral employment of RSEs were independent of PhD qualifications.

The actual sectoral employment data set contains more detailed classification on the type of qualification for an RSE. Table 12.5 cross-classifies the sectoral employment of RSEs against three qualification levels (bachelor's, master's or PhD). Here there is a distinct ordering in the level of qualification. As discussed earlier, such ordinal information can be incorporated into the analysis to achieve higher power in the assessment of whether the probability of employment as an RSE in the private sector is dependent on the qualification level (highest for PhD and lowest for bachelor's).

Prior to the use of rigorous statistical hypothesis testing procedures, a preliminary analysis can be conducted by just looking at the data shown in Table 12.5, which includes the percentage of RSEs employed in each sector together with the usual statistics ignoring the ordinal information. From the data, there appears to be a decreasing trend in the percentage of RSEs employed in the private sector at each higher qualification level. The adjusted residuals in Table 12.5 also show a similar trend. The adjusted residuals for the number of RSEs employed in the private sectors are both negative at higher qualification levels corresponding the master's and PhD degrees.

Table 12.4 Computed statistics for comparison of sectoral employment of PhD RSEs.

	Employment in			
	Private sector	Public sector	p_{i+}	Total
RSE with PhD	781	3 282	0.215	4 063
X^2	1 171	1 850		
R_{adj}	−62	62		
RSE without PhD	10 815	4 057	0.785	14 872
X^2	320	505		
R_{adj}	62	−62		
p_{+j}	0.613	0.387		
Total	11 596	7 339		

Table 12.5 Statistics on sectoral employment of RSEs by level of qualifications.

Level of Qualifications of RSEs	Employment in		Percentage employed in private sector as RSE	Total
	Private Sector	Public Sector		
PhD	781	3 282	21.5%	4 063
X^2	1 171	1 851		
R_{adj}	−62	62		
Master's	2 831	2 073	25.9%	4 904
X^2	10	16		
R_{adj}	−6	6		
Bachelor's	7 984	1 984	52.6%	9 968
X^2	579	914		
R_{adj}	56	−56		
Percentage of RSEs in each sector	61.2%	38.8%		
Total	11 596	7 339		18 935

Both sample percentages and adjusted residuals show that there is a possible negative correlation between qualification level and probability of employment in the private sector as an RSE.

As the M^2 statistic is more sensitive to departures from the null hypothesis when ordinal information in the categorical data is accounted for, it is used here to statistically assess the presence of relationships between the response and explanatory variables. The ordinal M^2 statistic requires scores for each level of the variables. In this case, arbitrary equally spaced scores are assumed for each level of the variables. For the explanatory variables based on the level of qualification, a score of $v_1 = 1$ is assumed for RSEs with Bachelors qualifications, $v_2 = 2$ for RSEs with Masters qualifications and $v_3 = 3$ for RSEs with PhD qualifications. Similarly, arbitrary equally spaced scores are assumed for the two types of sectoral employment. Since the response variables only have two levels, RSEs in private sector employment are assumed to be represented by a score of $u_1 = 1$ and those in public sector employment are assumed to have a score of $u_1 = 0$. Using this scoring system, the sample Pearson product-moment correlation, r, is found to be -0.485 and the corresponding M^2 statistic is found to be 4446. Since this is very much higher than the critical value of the χ^2 statistic having 1 degree of freedom at any reasonable significance level, the null hypothesis of independence is rejected with very strong evidence.

12.3.1 Sample proportions, relative risks, and odds ratio

Further statistical inference can be made on the degree of association between binary variables in two-way contingency tables by comparing differences in the proportions of total counts falling in each cell. Apart from this proportion of counts, two other useful measures of association for two-way contingency tables are the relative risk

and odds ratio. The use of these measures is illustrated using the sectoral employment example.

Consider the contingency table shown in Table 12.3. Using the generic terms 'success' to represent employment in the private sector and 'failure' to refer as employment in the public sector (these labels can of course be interchanged), π_{11} is defined as the proportion of successes for RSEs with PhDs and π_{21} is defined as the proportion of successes for RSE without PhDs. The difference between these proportions essentially compares the success probabilities (or the probability of being employed in the private sector as an RSE for this example). A more formal way to define the difference between these probabilities using conditional probabilities is:

$$P(\text{Employed in Private Sector as RSE}|\text{PhD Qualification}) -$$

$$P(\text{Employed in Private Sector as RSE}|\text{No PhD Qualification}).$$

Assuming that the counts in both rows follow independent binomial distributions and using the sample proportions, p_{11} and p_{21}, as estimates for the success probabilities, the sample difference $(p_{11} - p_{21})$ estimates the population difference with the estimated standard error given by

$$se(p_{11} - p_{21}) = \sqrt{\frac{p_{11}(1 - p_{11})}{n_{1+}} + \frac{p_{21}(1 - p_{21})}{n_{2+}}}.$$

Thus, a 95% confidence interval for the sample difference is

$$(p_{11} - p_{21}) \pm z_{\alpha/2} se(p_{11} - p_{21}).$$

The sample difference for the contingency table shown in Table 12.3 is -0.53 and its corresponding confidence interval is $(-0.55, -0.52)$. Since the interval contains only negative values, it can be concluded that the PhD qualification has a negative influence on the probability of successfully finding employment in the private sector. Apart from the difference in the conditional probability of being employed in the private sectors, conditional probabilities of employment in the public sector can also be elicited from the contingency table. The difference between conditional probabilities for employment in the public sector as an RSE is 0.53 with a corresponding confidence interval of (0.52, 0.55). This difference is formally expressed as

$$P(\text{Employed in Public Sector as RSE}|\text{PhD Qualification}) -$$

$$P(\text{Employed in Public Sector as RSE}|\text{no - PhD Qualification}).$$

Relative risk is a more elaborate measure than the difference between proportions of successes. Rather than considering the absolute differences of proportions $(\pi_{11} - \pi_{21})$, it captures the relative differences are through a ratio of proportions (π_{11}/π_{21}). Utilizing this measure, bias due to scaling effect inherent in absolute differences is mitigated. The sample estimate of the relative risk is p_{11}/p_{21}. As it is a ratio of two random variables, its sampling distribution can be highly skewed and its confidence interval relatively more complex. Both the relative risk and its corresponding confidence intervals can be evaluated with software such as MINITAB. The sample estimates of four

Table 12.6 Relative risks and 95% confidence intervals

Conditional probability	Relative risk	95% confidence interval
$\frac{P(\text{Employed in Private Sector as RSE}\vert\text{no - PhD Qualification})}{P(\text{Employed in Private Sector as RSE}\vert\text{PhD Qualification})}$	3.78	(3.55, 4.03)
$\frac{P(\text{Employed in Private Sector as RSE}\vert\text{PhD Qualification})}{P(\text{Employed in Private Sector as RSE}\vert\text{no - PhD Qualification})}$	0.26	(0.25, 0.28)
$\frac{P(\text{Employed in Public Sector as RSE}\vert\text{no - PhD Qualification})}{P(\text{Employed in Public Sector as RSE}\vert\text{PhD Qualification})}$	0.34	(0.33, 0.35)
$\frac{P(\text{Employed in Public Sector as RSE}\vert\text{PhD Qualification})}{P(\text{Employed in Public Sector as RSE}\vert\text{no - PhD Qualification})}$	2.96	(2.87, 3.05)

possible relative risk measures are shown in Table 12.6, together with their confidence intervals. Since a relative risk measure of 1 indicates that there are no differences in the probabilities conditional upon the qualification levels, it can be observed that the probability of finding employment in the private sector is 2.78 times (=3.78 − 1) higher without PhD qualification. Assuming the frequency counts follow independent binomial distributions for both categories of the explanatory variable, the confidence levels indicate that the probability of finding employment in the private sector is at least 2.55 times (=3.55 − 1) higher if the RSE does not have a PhD.

Another commonly used measure of association for contingency tables is the odds ratio. In contrast to relative risk which is a ratio of two probabilities, odds ratio is a ratio of two odds. Given that an explanatory variable for a two-way contingency table has I levels with i representing the ith level of the explanatory variable, the odds of a success for the ith category is defined as

$$odds_i = \frac{\pi_{i1}}{1 - \pi_{i1}}. \tag{12.10}$$

It can be observed that the odds are essentially the ratio of successes to failures for each category of the explanatory variable. The odds ratio is then defined as the ratio of these odds:

$$\theta = \frac{odds_1}{odds_2} = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}. \tag{12.11}$$

The sample estimate of the odds ratio can be expressed using the frequency counts in each cell:

$$\hat{\theta} = \frac{odds_1}{odds_2} = \frac{n_{11}n_{12}}{n_{12}n_{21}}. \tag{12.12}$$

For the example on the sectoral employment of RSEs by qualification levels with data shown in Table 12.3, using the same definition of 'success' for each level of the explanatory variable (with and without PhD qualifications), the sample odds of a

PhD RSE being employed in the private sector are 0.23 and the corresponding sample odds of a non-PhD RSE being employed in the private sector are 2.67. The sample odds ratio, defined as the odds of success for non-PhD RSEs over the odds of success for PhD RSEs, is 11.2. The confidence interval for this odds ratio can be found by using a large-sample normal approximation to the sampling distribution of $\ln \hat{\theta}$. The mean of this distribution is $\ln \hat{\theta}$ with the asymptotic standard error given by

$$ASE(\ln \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}.$$

The confidence interval can thus be evaluated from

$$\ln\left(\hat{\theta}\right) \pm z_{\alpha/2} ASE\left(\ln\left(\hat{\theta}\right)\right).$$

The confidence interval of the odds ratio for this example is evaluated to be (10.3, 12.2).

12.4 LOGISTIC REGRESSION APPROACH

The previous section treats statistical inference procedures for detecting the presence of relationships between the response and explanatory variables. These techniques essentially form the bedrock of statistical tools for categorical data analysis. In this section a class of model-based statistical approaches based on the logistic regression model for categorical data analysis is introduced.

There are many benefits associated with model-based approaches to characterizing the relationships between response and explanatory variables. Appropriate models allow statistically efficient estimation of the strength and importance of the effect of each explanatory variable and the interactions between them. Model-based techniques generally allow for more precise statistical estimates and stronger statistical inferences. Furthermore, a model-based paradigm is able to handle more complex cases involving multiple explanatory variables. In this section, a brief discussion of the logistic regression model as a special case of the generalized linear model is presented. This is followed by a description of logistic regression for the case of a single explanatory variable. An introduction to handling categorical responses with multiple explanatory variables is then presented.

12.4.1 Logit link for logistic regression

Logistic regression models are essentially generalized linear models (GLMs) which characterize relationships between a binary response variable and explanatory variables via a logit link function,

$$\ln\left(\frac{\pi\left(x\right)}{1-\pi\left(x\right)}\right)$$

where $\pi(x)$ is the probability of success for the binary variable.

A link function in generalized linear modeling terminology essentially describes the functional relationship between the random and systematic component in a GLM.

The random component in GLMs describes the distribution of the random response observations. The systematic component characterizes the relationship of the expected response with the explanatory variables through the link function. In GLMs, the systematic component involves a function which is linear in the parameters. Using a simple linear model with one explanatory variable, x, the GLM for response variables in logistic regression has the form

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x, \tag{12.13}$$

where α is a constant, and β is a slope parameter for the explanatory variable, x.

Besides the logit link, the relationship between the random component and the systematic component can be modeled with other link functions such as the Gompert, $\ln(-\ln(1-\pi))$, and the probit/normit, $\Phi^{-1}(\pi)$, where $\Phi(\cdot)$ is the normal cumulative distribution function. In practice, the canonical link based on the response distribution is most commonly used in generalized linear modeling. The canonical link is the link function which uses the natural parameter of the mean of the response distribution. Binary response variables can be modeled with a Bernoulli distribution having the probability of success as its expected value or a binomial distribution when the response is a sum of such Bernoulli distributed binary responses. The canonical link function for the random variables which follows either the Bernoulli or binomial distribution is the logit link.

The use of the logit link in logistic regression models offers many other distinct advantages for modeling binary response variables. The logit link function essentially depicts the odds of 'success'. The odds of success can be evaluated directly by taking the antilog of the GLM with the logit link function:

$$\frac{\pi(x)}{1-\pi(x)} = e^{\alpha}e^{\beta x} \tag{12.14}$$

From (12.14) it can be observed that the odds change multiplicatively by a factor of e^{β} with each unit increase in x. The modeling of log odds also implies that the logistic regression model can be readily used for the analysis of data from retrospective sampling designs[1] through the use of odds ratio. As can be observed from the definition of the odds ratio in (12.11) and (12.12), the ratio does not change when the response and explanatory variables are interchanged. In retrospective sampling designs such as case–control studies, the number of cases for each category of the response variables is fixed by the sampling design. Hence, in order to evaluate the conditional distributions of the response, the symmetry property of the odds ratio can be used.

12.4.2 Logistic regression with a single continuous explanatory variable

Agresti gives an example on the study of the nesting of horseshoe crabs.[1] Each female horseshoe crab in the study has a male crab attached to her which is considered her 'husband'. An investigation was conducted to determine the factors which affect whether a female horseshoe crab has any other male crabs residing nearby apart from her husband. These male horseshoe crabs are called 'satellites'. The presence of satellites is thought to depend on various factors. Some of these possible factors are the female crabs' color, spine condition, weight, and carapace width. As the main

Table 12.7 Characteristics of female horseshoe crabs.

Color (x_C)	Spine condition (x_S)	Carapace width (x_{Width})	Weight (x_{Weight})	Response* (y)	Color (x_C)	Spine condition (x_S)	Carapace width (x_{Width})	weight (x_{Weight})	Response (y)
3	3	28.3	3050	1	3	3	28.7	3150	1
4	3	22.5	1550	0	3	1	26.8	2700	1
2	1	26	2300	1	5	3	27.5	2600	0
4	3	24.8	2100	0	3	3	24.9	2100	0
4	3	26	2600	1	2	1	29.3	3200	1
3	3	23.8	2100	0	2	3	25.8	2600	0
2	1	26.5	2350	0	3	2	25.7	2000	0
4	2	24.7	1900	0	3	1	25.7	2000	1
3	1	23.7	1950	0	3	1	26.7	2700	1
4	3	25.6	2150	0	5	3	23.7	1850	0
4	3	24.3	2150	0	3	3	26.8	2650	0
3	3	25.8	2650	0	3	3	27.5	3150	1
3	3	28.2	3050	1	5	3	23.4	1900	0
5	2	21	1850	0	3	3	27.9	2800	1
3	1	26	2300	1	4	3	27.5	3100	1
2	1	27.1	2950	1	2	1	26.1	2800	1
3	3	25.2	2000	1	2	1	27.7	2500	1
3	3	29	3000	1	3	1	30	3300	1
5	3	24.7	2200	0	4	1	28.5	3250	1
3	3	27.4	2700	1	4	3	28.9	2800	1
3	2	23.2	1950	1	3	3	28.2	2600	1
2	2	25	2300	1	3	3	25	2100	1
3	1	22.5	1600	1	3	3	28.5	3000	1
4	3	26.7	2600	1	3	1	30.3	3600	1
5	3	25.8	2000	1	5	3	26.2	1300	0

*1, satellites present; 2, satellites absent.

objective of the analysis is to predict whether there are any satellites present, the response, y, is assumed to be a binary variable of the following form:

$$y = \begin{cases} 0, & \text{if satellites present,} \\ 1, & \text{if satellites absent.} \end{cases}$$

Part of the original data set is reproduced in Table 12.7. The analysis in this example is conducted based on this partial data set.

In an initial analysis, a single explanatory variable based on the carapace width is assumed. Since the response is binary and can be assumed to follow a Bernoulli distribution, a simple logistic regression model with a single explanatory variable of the following form is postulated:

$$\ln\left(\frac{\pi_{\text{Pres}}(x_{\text{Width}})}{1-\pi_{\text{Pres}}(x_{\text{Width}})}\right) = \alpha + \beta x_{\text{Width}} \tag{12.15}$$

Table 12.8 Logistic regression table for simple model with one explanatory variable.

						95% Confidence Limits	
Predictor	Coefficient	SE Coef*	z-statistic	p-value	Odds ratio	Lower	Upper
Constant	$\hat{\alpha} = -19.43$	6.09	−3.19	0.001			
x_{Width}	$\hat{\beta} = 0.77$	0.23	3.25	0.001	2.15	1.36	3.42

*Standard error of coefficient estimates.

where x_{Width} is the carapace width of a female horseshoe crab and $\pi_{\text{Pres}}(\cdot)$ is the probability of finding a satellite crab nearby.

MINITAB offers the facility to analyze such simple logistic regression models and the output from this analysis is reproduced in Table 12.8. The maximum likelihood estimates (MLEs) of the coefficients are given in the output as $\hat{\alpha} = -19.43$ and $\hat{\beta} = 0.77$. From Table 12.8, these coefficient estimates are found to be significant by looking at the z statistics and p-values. The z statistics of the estimates are evaluated from:

$$z_\alpha = \frac{\hat{\alpha}}{ASE\,(\hat{\alpha})} \quad \text{and} \quad z_\beta = \frac{\hat{\beta}}{ASE\,(\hat{\beta})}$$

where, $ASE\,(\hat{\alpha})$ is the asymptotic standard error of the MLE of parameter $\hat{\alpha}$, and similarly for $\hat{\beta}$.

The hypothesis test conducted for these estimates is based on an asymptotic normal distribution at the 5% level of significance. The formal hypothesis test for the significance of any parameter γ in the simple logistic regression model is stated as follows:

$$H_0\text{: } \gamma = 0 \quad \text{vs.} \quad H_1\text{: } \gamma \neq 0$$

Such a hypothesis test can also be conducted via the z^2 statistic (or Wald statistic) which follows a large sample χ^2 distribution with 1 degree of freedom. Alternatively, a likelihood ratio test can be conducted. The likelihood ratio test utilizes the ratio of two maximum log likelihood functions. The two maximum log likelihood functions are evaluated based on the maximized log likelihoods of a *reduced* model without any explanatory variables (l_0) and the log likelihood of a *full* model (l_1) postulated by the alternative hypothesis. For a simple logistic regression model with only two parameters, α and β, l_0 denotes the maximized log likelihood when $\beta = 0$ and l_1 denotes the maximized likelihood with both α and β parameters present in the model. The likelihood ratio statistic can be expressed as

$$G^2 = -2\ln\left(\frac{l_0}{l_1}\right). \tag{12.16}$$

Under the null hypothesis, the likelihood ratio statistic for the logistic regression model with one explanatory variable follows a large sample χ^2 distribution with $p - 1$ degrees of freedom, where p is the number of parameters estimated. For the horseshoe crab study, the maximized log likelihoods for the *full* and *reduced* models are −24.618 and −33.203, respectively. This gives a G^2 statistic of 17.170. This is much

larger than the critical value of χ^2 with 1 degree of freedom at the 5% significance level. Hence the null hypothesis is rejected, implying that the explanatory variable has significant effects on the response.

The odds ratio given in Table 12.8 as evaluated in MINITAB essentially gives the rate of increase in the odds given a unit change in the explanatory variable. For the horseshoe crab example, the rate of increase in the odds of finding satellites given a unit increase in the width of the female horseshoe crab is 2.15. As shown in equation (12.14), this rate of increase in odds is given by e^{β}. The confidence interval for the odds ratio can thus be evaluated from

$$e^{\hat{\beta} \pm z_{\alpha/2} ASE(\hat{\beta})}.$$

With the parameter estimates, the probabilities of the presence of satellites given the carapace width of the female crab can also be evaluated as:

$$\hat{\pi}_{\text{Presence}}(x_{\text{Width}}) = \frac{\exp(\hat{\alpha} + \hat{\beta} x_{\text{Width}})}{1 + \exp(\hat{\alpha} + \hat{\beta} x_{\text{Width}})}$$

These are known as the *event probabilities*. These event probabilities are plotted against the explanatory variables in Figure 12.1.

After fitting the logistic regression model, it is important to examine the suitability of the model in light of the prediction generated. The different techniques available for such tests are commonly referred to as goodness-of-fit (GOF) tests as they compare the quality of fit between the predicted responses and the actual responses. The most common GOF tests in categorical data analysis are those based on likelihood ratios and Pearson residuals. Here, GOF tests based on likelihood ratios are discussed.

The motivation for likelihood ratio based GOF tests basically arises from the maximum likelihood theory underlying the parameter estimation process. A typical

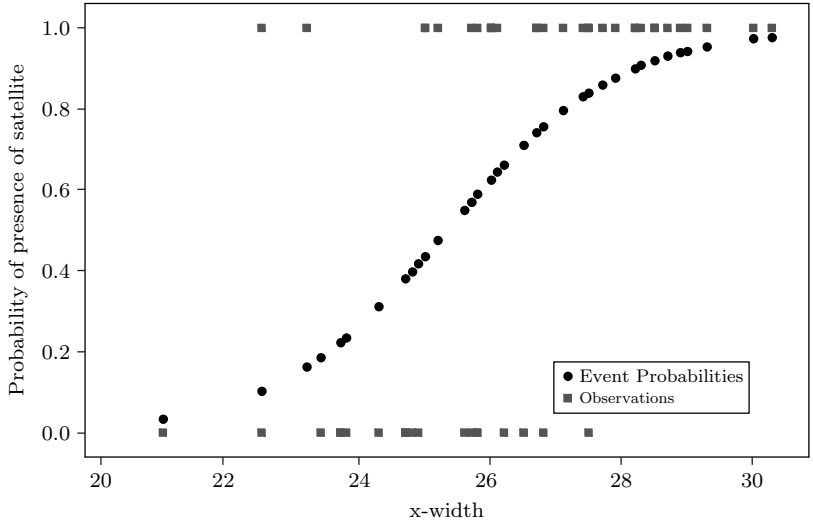


Figure 12.1 Plot of event probability against explanatory variable, x_{Width}, for the horseshoe crab example.

likelihood ratio statistic is given by equation (12.16). The ratio of two likelihoods can be written as a difference between 2 log likelihoods. A special case of the likelihood ratio written as a difference of two log likelihoods which is used in likelihood ratio GOF tests is the 'deviance' statistic given by

$$\text{Deviance} = -2\,(L_{\text{N}} - L_{\text{S}})$$

where L_{S} is the maximum log likelihood for a saturated model (with a separate parameter for each observation, resulting in a perfect fit of the observed data) and L_{N} is the maximum log likelihood for a non-saturated model. For logistic regression involving binomial random variables, the deviance statistic has an asymptotic null χ^2 distribution with $N - p$ degrees of freedom, where N is the number of distinct observations and p is the number of parameters in the model. By utilizing the χ^2 distribution, a statistical hypothesis test can be conducted to validate the explanatory power of a postulated non-saturated model compared to the perfect explanatory power of a saturated model for a given data set. The null hypothesis for this test is that all parameters that are in the saturated model but not in the postulated non-saturated model are zero. For the horseshoe crab example with a single continuous explanatory variable (carapace width) the deviance of this model is calculated as 33.28 with 33 degrees of freedom. This gives a p-value of 0.454. Hence, the null hypothesis cannot be rejected at the 5% significance level.

It should be noted that there are other GOF tests for categorical data analysis with continuous explanatory variables such as the Hosmer–Lemeshow tests[2] and those based on Pearson residuals. Such tests can be easily evaluated with statistical software such as MINITAB. Details of these tests are given by Agresti.[1] The p-value is evaluated together with the GOF statistics in MINITAB and can be used to provide a more comprehensive judgment on the GOF of the estimated model.

12.4.3 Logistic regression with single categorical explanatory variable

Logistic regression can easily be extended to model categorical explanatory variables via the use of dummy variables. The procedure for handling categorical explanatory variables is described in this section using the above horseshoe crab example. The explanatory variable of interest is taken to be the spine condition. From Table 12.7, it can be observed that spine condition is essentially a categorical variable with three levels. Two binary dummy variables which take values of 0 or 1 are necessary to fully describe the three levels of spine condition. The two dummy variables are denoted as c_{S1} and c_{S2}. The logistic regression model with a single categorical explanatory variable is given by

$$\text{In}\left(\frac{\pi_{\text{Pres}}\,(c_{S1}, c_{S2})}{1 - \pi_{\text{Pres}}\,(c_{S1}, c_{S2})}\right) = \alpha_S + \beta_{S1}c_{S1} + \beta_{S2}c_{S2}.$$

MINITAB offers the facility to analyze logistic regression models with categorical explanatory variables. The output from MINITAB using the horseshoe crab dataset is reproduced in Table 12.9. From the output, the MLE coefficients of both the constant and the parameters of the two dummy variables, $\hat{\alpha}_S$, $\hat{\beta}_{S1}$ and $\hat{\beta}_{S2}$, are found to be significant at the 5% level by looking at the z-statistic and p-value. Similar to the logistic regression model with a single continuous explanatory variable, the *p-value* is based on an asymptotic normal distribution. The logit function for the spine conditions

Table 12.9 Logistic regression table for model with single categorical explanatory variable.

Predictors			Coefficient	SE Coeff.	Z	p-value	Odds ratio	95% CI of odds ratio	
								Lower	Upper
Constant			$\hat{\alpha}_S = 1.87$	0.76	2.46	0.014			
Spine Condition	Dummy variable combinations								
	c_{S1}	c_{S2}							
2	1	0	$\hat{\beta}_{S1} = -2.28$	1.19	−1.92	0.055	0.10	0.01	1.05
3	0	1	$\hat{\beta}_{S2} = -1.74$	0.84	−2.06	0.039	0.18	0.03	0.92

1, 2 and 3 can thus be evaluated by considering the three combinations of (c_{S1}, c_{S2}) at (0,0), (1,0) and (0,1) respectively. The linear functions for the log odds evaluated for each spine condition are listed in Table 12.10.

The odds ratios evaluated at each setting of the categorical variable for spine conditions 2 and 3 shown in Table 12.9 are odds ratios in comparison with the case of the horseshoe crab having spine condition *1*. The odds ratio is evaluated by taking the antilog of the logit functions shown in Table 12.10. The odds ratios corresponding to spine condition 2 and 3 are respectively

$$\exp\left[\left(\hat{\alpha}_S + \hat{\beta}_{S1}\right) - \hat{\alpha}\right] \text{and} \exp\left[\left(\hat{\alpha}_S + \hat{\beta}_{S2}\right) - \hat{\alpha}_S\right].$$

These odds ratios compare the relative differences in odds of finding satellite crabs for female horseshoe crabs with different spine conditions against a reference spine condition, which is spine condition 1 in this case. The confidence intervals for these odds ratio can be approximated using the asymptotic standard error of the parameter estimates. Using MINITAB, these confidence intervals are evaluated as (0.01, 1.05) and (0.03, 0.92) for the odds ratio spine conditions 2 and 3 respectively. From these confidence intervals, there appears to be no significant difference in the odds of finding satellite crabs for female horseshoe crabs having spine conditions 1 and 2. However, the estimated odds of finding satellite crabs for female horseshoe crabs with spine condition 3 appears to be at most 0.92 times different from that of horseshoe crabs with spine condition 1.

Table 12.10 Logit funtions relevant to each spine condition.

Spine Condition	Dummy variable combinations		$\log\left(\frac{\pi_{\text{Pres}}(c_{S1}, c_{S2})}{1-\pi_{\text{Pres}}(c_{S1}, c_{S2})}\right)$
	c_{S1}	c_{S2}	
1	0	0	$\hat{\alpha}_S$
2	1	0	$\hat{\alpha}_S + \hat{\beta}_{S1}$
3	0	1	$\hat{\alpha}_S + \hat{\beta}_{S2}$

The preceding observations are based on assumption that all other possible explanatory variables are kept constant. As observed in Table 12.7, there are most probably other factors which simultaneously affect the presence of satellite crabs. More complex logistic regression models that can take into account multiple explanatory variables may be necessary to fully describe a model with higher predictive power. The plausibility of such models is investigated in the next subsection.

12.4.4 Multiple logistic regression

The ability of simple logistic regression models with a single explanatory variable to generalize to complex multiple logistic regression models with multiple variables is similar to that of the generalization of simple linear regression models to multiple linear regression models in OLS regression. A typical multiple logistic regression model for binary responses with k categorical variables and l continuous variables can be represented as follows:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_{C1}c_1 + \beta_{C2}c_2 + \cdots + \beta_k c_k + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_l x_l,$$

where α is a constant, c_i is the ith categorical explanatory variable and β_{Ci} its slope parameter, x_i is the ith continuous explanatory variable and β_i its slope parameter, and π is the probability of success.

In the horseshoe crab data shown in Table 12.7, the possible categorical variables are the crab color and spine condition. The continuous explanatory variables are the weight of crab and carapace width. The fitted model could potentially be of the following form:

$$\ln\left(\frac{\pi_{\text{Pres}}}{1-\pi_{\text{Pres}}}\right) = \alpha + \beta_{C1}c_1 + \beta_{C2}c_2 + \beta_1 x_1 + \beta_2 x_2 \tag{12.17}$$

where, c_1 is the color variable, c_2 the spine condition, x_1 the width, x_2 the weight, and π_{Pres} the probability of finding a satellite crab nearby. The categorical variable c_1 has four levels whereas c_2 has three; hence, three dummy variables are necessary to completely describe c_1 and two are needed for c_2. The three dummy variables for c_1 are represented by c_{1i} for $i = \{1, 2, 3\}$ and for c_2 are represented by c_{2j} for $j = \{1, 2\}$. MLEs of the parameters evaluated using MINITAB is shown in Table 12.11.

From Table 12.11, there appear to be no explanatory variables which are significant. This contradicts the earlier analysis with single categorical and continuous variables. Furthermore, the G^2 statistic calculated with Equation (12.16) using MINITAB is 26.3. The null hypothesis for the test based on this statistic states that the response is jointly independent of all the explanatory variables. Based on an asymptotic null χ^2 distribution with 7 degrees of freedom, there is very strong evidence of the presence of significant effects in at least one of the explanatory variables on the response. The fact that all effects in the logistic regression table in Table 12.11 showed up as insignificant instead could be indicative of the presence of significant multicollinearity between the explanatory variables. Multicollinearity essentially refers to the presence of significant relationships between the explanatory variables such that some or all of the

Table 12.11 Logistic regression table for multiple logistic regression model in Equation (17).

Predictors				Coefficient	SE Coeff.	Z	p-value	Odds ratio	95% CI of odds ratio	
									Lower	Upper
Constant				−16.095	8.683	−1.85	0.064			
	Dummy variable combinations									
Color	c_{11}	c_{12}	c_{13}							
3	1	0	0	1.393	1.277	1.09	0.275	4.03	0.33	49.22
4	0	1	0	0.692	1.491	0.46	0.643	2.00	0.11	37.08
5	0	0	1	−0.090	1.899	−0.05	0.962	0.91	0.02	37.80
	Dummy variable combinations									
Spine Condition	c_{21}		c_{22}							
2	1		0	−0.838	1.386	−0.60	0.546	0.43	0.03	6.55
3	0		1	−2.108	1.243	−1.69	0.090	0.12	0.01	1.39
Width				0.521	0.451	1.15	0.249	1.68	0.69	4.08
Weight				0.002	0.002	0.83	0.406	1.00	1.00	1.01

explanatory variables can be written in terms of the other explanatory variables. The presence of such multicollinearity will seriously affect both the parameter estimates and their corresponding variance estimates. Hence, some significant explanatory variables may be incorrectly rejected through the usual statistical hypothesis tests.

A simple and effective way to check for the presence of multicollinearity is through the use of scatter plots. The matrix plot function in MINITAB which places scatterplots depicting relationships between any two variables in a matrix form can be used to generate a visually effective summary of the scatterplots. Such a matrix plot is shown in Figure 12.2. From the matrix plot, it can be observed that there appears to be a significant relationship between the two continuous variables, weight and carapace width. In order to remedy the multicollinearity, the weight variable is removed from consideration. The regression is redone on the remaining variables. Table 12.12 shows the logistic regression table for this reduced model.

With the risk of multicollinearity mitigated, the potential for further refining the multiple regression model is investigated. The refinement process for complex regression models to obtain more parsimonious models that are as effective is generally known as *model selection*. The key goal in model selection is to achieve a parsimonious model that can fit the actual data well. In order to achieve this, the model has to balance the conflicting objective of achieving sufficient complexity to model the actual data well, yet sufficient simplicity for practical interpretation. A wide variety of model selection algorithms for GLMs incorporating different statistical criteria

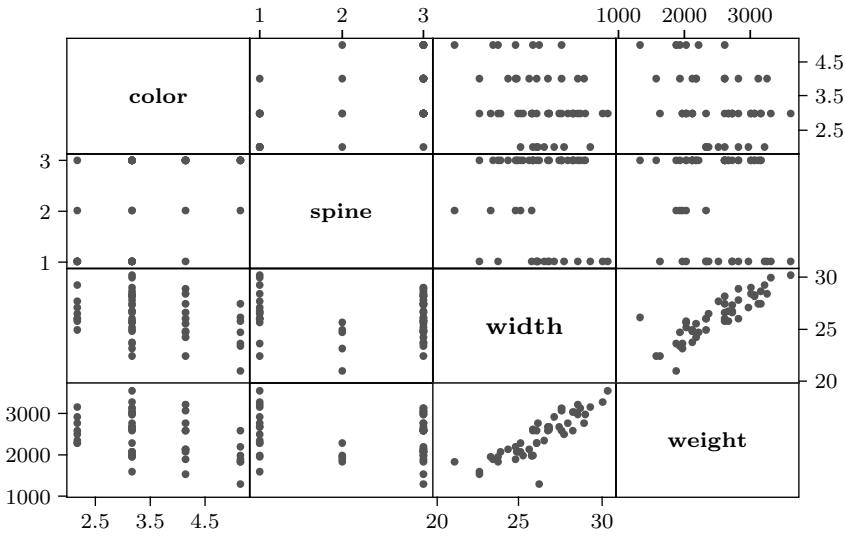


Figure 12.2 Matrix plot to check for multicollinearity due to significant correlations between explanatory variables.

Table 12.12 Logistic regression table for multiple logistic regression model.

Predictors				Coefficient	SE Coeff.	Z	p-value	Odds ratio	95% CI of odds ratio	
									Lower	Upper
Constant				−20.332	7.429	−2.74	0.006			
	Dummy variable combinations									
Color	c_{11}	c_{12}	c_{13}							
3	1	0	0	1.146	1.244	0.92	0.357	3.15	0.27	36.03
4	0	1	0	0.427	1.440	0.30	0.767	1.53	0.09	25.78
5	0	0	1	−0.841	1.737	−0.48	0.628	0.43	0.01	12.98
	Dummy variable combinations									
Spine Condition	c_{21}	c_{22}								
2	1	0		−0.849	1.413	−0.60	0.548	0.43	0.03	6.83
3	0	1		−1.893	1.191	−1.69	0.112	0.15	0.01	1.56
Width				0.833	0.284	1.15	0.003	2.30	1.32	4.01

have been proposed in literature. Algorithms that are similar to backward elimination and forward selection procedures in ordinary linear regression have been proposed. Criteria used in these algorithms include statistics such as the Akaike information criterion (AIC) and others based on the difference in likelihood ratios. In this chapter, we introduce a very simplistic backward elimination process which utilizes the difference in the likelihood ratios of two possible models. For model selection procedures using such measures, one of the models in the comparison must be a special case of the other. Assuming the likelihood ratio evaluated with equation (12.16) of the more complex model is given by G_1 and that of the simpler model is given by G_2, a hypothesis test can be conducted on the difference between these likelihood ratios, $G_D = G_1 - G_2$. Under the assumed models and specified hypothesis test, these differences typically follow an approximate null χ^2 distribution with degrees of freedom equivalent to the difference in degrees of freedom associated with the likelihood ratios of each model compared. Any significant difference indicates the need for additional variables in achieving a more predictive model. On the other hand, if there are no significant differences, the additional complexity that accompanies the more complex model is deemed unnecessary. Such a likelihood ratio statistic can implemented in a backward elimination procedure. The horseshoe crab example is used to demonstrate this backward elimination procedure.

In this example, the starting basis is assumed to be a model which considers only the main effects postulated. Strictly speaking, the starting model should include all the possible interaction effects together with these main effects in a backward elimination process of model selection. Nonetheless, for simplicity, the selection process is described taking only the main effects into account. The differences, G_D, between the likelihood ratios G_1 and G_2 associated with each pair of models are shown in Table 12.13. These differences are assumed to follow an approximate null χ^2 distribution with degrees of freedom equivalent to the difference in *df's* associated with each of the two model under comparison. From Table 12.13, it can be observed that for the model to have sufficient predictive power, the width variable is necessary as its removal results in significant differences in the likelihood ratios. It also appears that a model consisting of either spine condition and width or color and width is sufficient for achieving a model that is parsimonious yet effective, as from the G_D statistic there appears to be no significant difference in likelihood ratio when the other categorical

Table 12.13 Comparison of possible logistic regression models with only main effects.

Model No.	Model	G (df)	Models compared	G_D(df)	p-value
1	$\beta_0 + \beta_{C11}c_{11} + \beta_{C12}c_{12} + \beta_{C13}c_{13} + \beta_{C21}c_{21} + \beta_{C22}c_{22} + \beta_1 x_1$	25.554 (6)			
2	$\beta_0 + \beta_{C11}c_{11} + \beta_{C12}c_{12} + \beta_{C13}c_{13} + \beta_{C21}c_{21} + \beta_{C22}c_{22}$	13.206 (5)	1–2	12.348 (1)	0.000
3	$\beta_0 + \beta_{C11}c_{11} + \beta_{C12}c_{12} + \beta_{C13}c_{13} + \beta_1 x_1$	22.559 (4)	1–3	2.995 (2)	0.223
4	$\beta_0 + \beta_{C21}c_{21} + \beta_{C22}c_{22} + \beta_1 x_1$	22.443 (3)	1–4	3.111 (3)	0.375
5	$\beta_0 + \beta_1 x_1$	17.170 (1)	4–5	5.273 (1)	0.02

Table 12.14 Comparison of possible logistic regression models with only main effects.

Predictors			Coefficient	SE Coeff.	Z	p-value	Odds ratio	95% CI of odds ratio Lower	Upper
Constant			−19.761	6.515	−3.03	0.002			
Spine condition	Dummy variable combinations c_{21}	c_{22}							
2	1	0	−0.962	1.397	−0.69	0.491	0.38	0.02	5.91
3	0	1	−2.097	1.055	−1.99	0.047	0.12	0.02	0.97
Width			0.837	0.256	3.27	0.001	2.31	1.40	3.82

variable is added. Here, a model with spine and carapace width is adopted. The logistic regression table for the refined logistic regression model with spine and carapace width is shown in Table 12.14.

From Table 12.14, it can be observed from the *p-value* for the coefficient estimates that the dummy variable representing spine condition 2 (p-value 0.491) appears to be unnecessary. Spine conditions 1 and 2 can potentially be grouped together. It is expected that this will not affect the predictive power of the simplified model. Hence, a simplified model is fitted with only one dummy variable to differentiate crabs with spine condition 3 from the others. The parameter MLEs together with their corresponding p-value and odds ratio are shown in Table 12.15. From equation (12.16), this model gives a likelihood ratio statistic of 21.964 with 2 degrees of freedom. Compared with the earlier model with an additional dummy variable distinguishing spine condition 2, the difference in the likelihood ratio statistic, G_D, is 0.479. The p-value based on an approximate null χ^2 distribution with 1 degree of freedom, is 0.5 implying that this simplified model is no different from the earlier model with 2 dummy variables differentiating spine conditions 1, 2 and 3.

Table 12.15 Comparison of possible logistic regression models with only main effects.

Predictors	Coefficient	SE Coeff.	Z	p-value	Odds ratio	95% CI of odds ratio Lower	Upper
Constant	−20.871	6.339	−3.29	0.001			
Spine Condition 3	−1.732	0.857	−2.02	0.043	0.18	0.03	0.95
Width	0.866	0.251	3.44	0.001	2.38	1.45	3.90

12.5 CONCLUSION

Key concepts behind categorical data analysis utilizing the contingency table and logistic regression approaches have been described in this chapter. These are the two most important families of categorical data analysis techniques and form the core ideas from which many other more sophisticated statistical techniques for categorical data analysis are developed. The use of categorical data analysis procedures is primarily driven by the presence of response variables which are not continuous. Such data have been observed to frequently arise in many Six Sigma applications especially in service or transactional processes. Categorical data analysis techniques are expected to become more relevant with the proliferation of Six Sigma in different industries.

REFERENCES

1. Agresti, A. (2002) *Categorical Data Analysis.* New York: John Wiley & Sons, Inc.
2. Hosmer, D. W. and Lemeshow, S. (1989) *Applied Logistic Regression.* New York: John Wiley & Sons, Inc.

13

A Graphical Approach to Obtaining Confidence Limits of C_{pk}

L. C. Tang, S. E. Than and B. W. Ang

The process capability index C_{pk} has been widely used as a process performance measure. In practice this index is estimated using sample data. Hence it is of great interest to obtain confidence limits for the actual index given a sample estimate. In this chapter we depict graphically the relationship between the process potential index (C_p), the process shift index (k) and the percentage nonconforming (p). Based on the monotone properties of the relationship, we derive two-sided confidence limits for $k$and C_{pk} under two different scenarios. These two limits are combined using the Bonferroni inequality to generate a third type of confidence limit. The performance of these limits of C_{pk} in terms of their coverage probability and average width is evaluated by simulation. The most suitable type of confidence limit for each specific range of k is then determined. The usage of these confidence limits is illustrated with examples. Finally, a performance comparison is done between the proposed confidence limits and three non-parametric bootstrap confidence limits. The results show that the proposed method consistently gives the smallest width and yet provides the intended coverage probability.

This chapter is based on the article by L. C. Tang, S. E. Than and B. W. Ang, 'A graphical approach to obtaining confidence limits of C_{pk}', *Quality and Reliability Engineering International*, 13(6), 1997, pp. 337–346, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

13.1 INTRODUCTION

The process capability indices C_p and C_{pk} are commonly used to determine whether a process is capable of operating within specification limits. Capability indices can be used to relate the process parameters μ and σ to engineering specifications that may include unilateral or bilateral tolerances with or without a target (or nominal) value. Kane[1] noted that these indices provide an easily understood language for quantifying the performance of a process. The capability measurement is compared with the tolerance in order to judge the adequacy of the process. According to Sullivan,[2] a process with high process capability index values would effectively eliminate inspection and defective material and therefore eliminate costs associated with inspection, scrap and rework.

Montgomery[3] presented some recommended guidelines for minimum values of C_p. If the estimated indices are larger than or equal to the respective recommended minimum values, then the process is claimed to be capable. For instance, if one desires a nonconforming (NC) proportion of 0.007 %, then a minimum value of $C_p = 1.33$ is recommended for an ongoing process, assuming that the process is entered. However, owing to sampling error, the estimated indices $\hat{C}_p$ and $\hat{C}_{pk}$ are likely to be different from the true indices C_p and C_{pk}. The recommended minimum values are for the true indices. Hence, even when the estimated index is larger than the minimum value, one may not be able to claim with a high level of confidence that the true index is indeed larger than the minimum value. This problem arises because the sampling error increases when the sample size used for estimation decreases. Thus a mere comparison between the estimated index and the recommended minimum values may not be a good process capability gauge, particularly when the sample size is small. This situation may arise in a short-run production where the initial samples available for process qualification are limited.

The above problem can be overcome by using lower confidence limits for C_{pk}. Unfortunately, it is not straightforward to obtain a confidence bound for C_{pk}. Difficulties in the construction of confidence limits for C_{pk} arise from the rather complicated way in which μ and σ appear in the expression for C_{pk}. Wang[4] pointed out that this complication has made the development of confidence limits for C_{pk} a much more difficult problem than simply combining the confidence limits of μ and σ to give a maximum and minimum C_{pk}.

Several methods for constructing lower confidence limits for C_{pk} have been presented in the literature. Reasonably comprehensive comparisons of these methods are given by Kushler and Hurley.[5] In that study, comparisons of the Chou approach,[6] the non-central t distribution approach and other approaches were made by numerical integration of the joint density function of a sample mean and a sample variance. The methods were then compared by their miss rates, that is, the frequency with which the $1 - \alpha$ confidence limits constructed by each method exclude true C_{pk} values. Kushler and Hurley concluded that the Chou approach is overly conservative. The non-central t distribution method, which is exact for a unilateral specification, gives a better approximation when the process is operating appreciably off-center.

Three normal approximations to the sampling distribution of $\hat{C}_{pk}$ are available in the literature. One involving gamma functions was derived by Zhang *et al.*[7] An alternative expression developed by Bissell[8] produces remarkably accurate lower $100(1 - \alpha)$ % confidence limits for C_{pk} for $n \geq 30$. The formula derived is relatively easy to compute

and is based on a Taylor series. The normal approximation presented by Kushler and Hurley[5] is a simplification of Bissell's result.

Another approach, also by Kushler and Hurley,[5] used the fact that since C_{pk} depends simultaneously on μ and σ, a joint confidence region for these two parameters can be used to obtain a confidence bound for C_{pk}. The minimum value of C_{pk} over the region is the lower confidence bound. One standard way to define a joint confidence region is to use contours of the likelihood function from the likelihood ratio test. Such confidence regions are asymptotically optimal and have also been found to provide good small-sample behavior in a variety of problems. However, determining the confidence bound by this method requires complex calculations.

Franklin and Wasserman[9] proposed a nonparametric but computer-intensive bootstrap estimation to develop three types of bootstrap confidence intervals for C_{pk}: the standard bootstrap (SB) confidence interval, the percentile bootstrap (PB) confidence interval and the bias-corrected percentile bootstrap (BCPB) confidence interval. They presented an initial study of the properties of these three bootstrap confidence intervals. The results indicated that some of the 90 % bootstrap intervals provided 90 % coverage when the process was normally distributed and provided 84–87 % when the process was chi-square distributed. Practitioners, however, have to bear in mind that the practical interpretation of the index C_{pk} is questionable when normality does not hold.

To choose a method for determining confidence limits for process capability indices, Kushler and Hurley[5] considered ease of use and method performance as guidelines. An additional guideline which has so far been overlooked is that the method for constructing a confidence interval for C_{pk} may depend on the nature of the process – for example, whether it is a short-run production process or a homogeneous batch process. Here we propose a scheme for determining the confidence interval that takes this guideline into consideration.

In the following section we depict graphically the relationship between C_p, k and p. Based on the monotone property of the relationship, we derive the confidence limits of k and construct a graphical tolerance box that relates the confidence limits of C_p, k and p. Two different types of two-sided confidence limits for C_{pk} are constructed by exploiting the fact that the sampling error in C_{pk} originates independently of the sampling error in $k$and C_p. These confidence limits are then combined using the Bonferroni inequality to give a third, conservative, type of two-sided confidence limit. The effectiveness of these confidence limits in terms of their coverage probability and average width is evaluated by simulation. By observing the performance of these confidence limits over different ranges of k, we propose an approximation method (AM) for choosing the appropriate confidence limits for C_{pk}. We also illustrate the usage of AM confidence limits via two examples. Finally, we compare the performance of the proposed AM confidence limits with that of the bootstrap confidence limits developed by Franklin and Wasserman.[9] A discussion on the simulation results and recommendations on the type of confidence limits for different scenarios are presented.

13.2 GRAPHING C_p, k AND p

Consider a measured characteristic X with lower and upper specification limits denoted by LSL and USL, respectively. Measured values of X that fall outside these

limits are termed 'nonconforming' (NC). An indirect measure of potential capability to meet the settings *LSL*, *X*, and *USL* is the process capability index

$$C_p = \frac{USL - LSL}{6\sigma}. \tag{13.1}$$

Here σ denotes the standard deviation of *X*and is estimated by

$$s = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2}, \tag{13.2}$$

where n is the sample size and

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j. \tag{13.3}$$

Denoting the midpoint of the specification range by $m = (USL + LSL)/2$, the shift index k is given by

$$k = \frac{|m - \mu|}{(USL - LSL)/2}. \tag{13.4}$$

The proportion of NC product can be estimated in term of C_p and k by

$$\hat{p} = \Phi[-3(1 + \hat{k})\hat{C}_p] + \Phi[-3(1 - \hat{k})\hat{C}_p]. \tag{13.5}$$

A contour plot of equation (13.5) on two different scales (Figures 13.1 and 13.2) reveals some interesting and useful properties. It can be seen from the figures that while (C_p, k) uniquely determine p, there also exists a unique value of k for each (C_p, p). Intuitively, we know that for each constant k, p increases as C_p decreases, and for each constant C_p, p increases as $k$increases. Figure 13.1 also reveals that for each

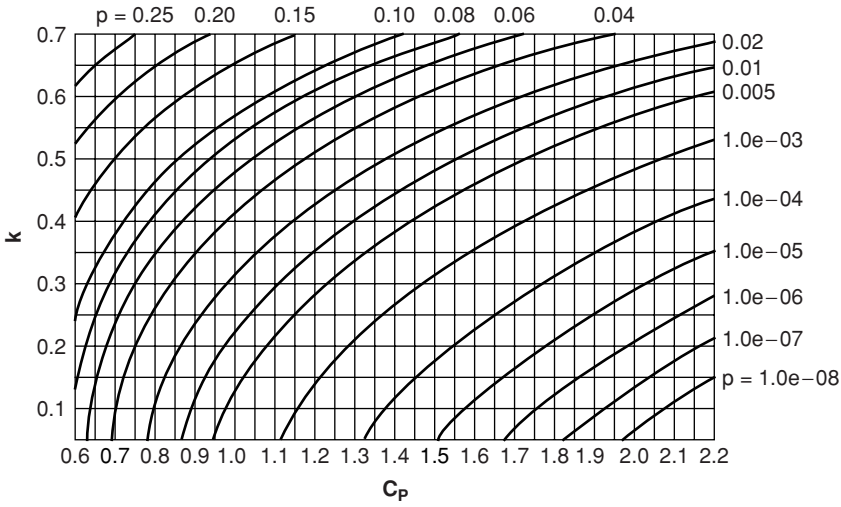


Figure 13.1 Behavior of C_p, k and p on linear scale.

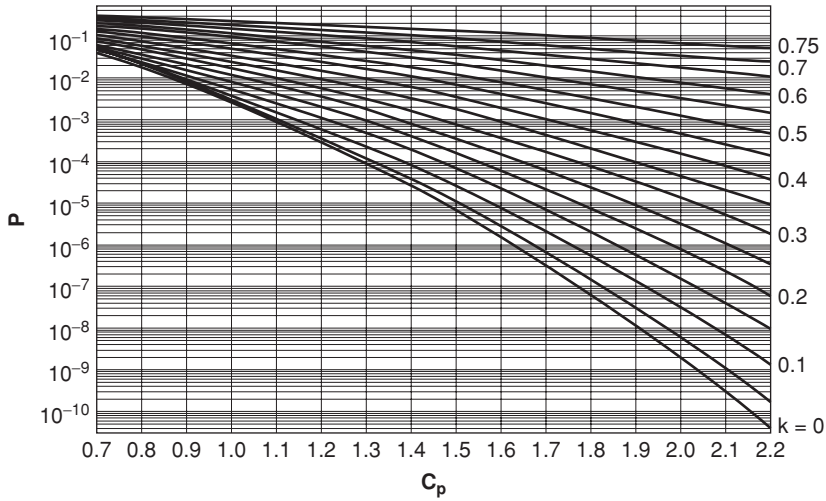


Figure 13.2 Behavior of C_p, k and p on log scale.

constant p, k is monotonic increasing with respect to C_p. This unique relationship between C_p, k and p allows us to derive the confidence limits of k from the easily computable confidence limits of C_p and p.

13.3 CONFIDENCE LIMITS FOR *k*

Consider a sample from a normally distributed stable process. A $100(1-\alpha)$ % confidence interval for C_p is

$$\left[\frac{USL-LSL}{6\hat{\sigma}}\frac{\chi^2_{n-1,\alpha/2}}{\sqrt{n-1}}, \frac{USL-LSL}{6\hat{\sigma}}\frac{\chi^2_{n-1,1-\alpha/2}}{\sqrt{n-1}}\right] \equiv \left[\frac{\chi^2_{n-1,\alpha/2}}{\sqrt{n-1}}\hat{C}_p, \frac{\chi^2_{n-1,1-\alpha/2}}{\sqrt{n-1}}\hat{C}_p\right], \quad (13.6)$$

where χ^2_{n-1} is the chi-square distribution value with $n-1$ degrees of freedom. From the same sample the maximum likelihood estimate of the fraction NC is given by $\hat{p}$, which is the number of NC units divided by n. The upper and lower confidence limits ($\bar{p}$ and $\underline{p}$) of $\hat{p}$ can be obtained from the binomial equation.

Treating k as the statistic of interest, we write

$$L(k : p, C_p) = p - \Phi\left[-3(1+k)C_p\right] + \Phi\left[-3(1-k)C_p\right]. \quad (13.7)$$

The unique and monotonic relationship between k and (C_p, p) permits the derivation of the upper and lower confidence limits of k, ($\bar{k}$, k), using ($\bar{C}_p$, $\bar{p}$) and ($\underline{C}_p$, $\underline{p}$), respectively. This is shown graphically in Figure 13.3 and can be expressed mathematically as

$$\underline{k} = \left\{k : L(k : p, C_p) = 0\right\}, \quad (13.8)$$

$$\bar{k} = \left\{k : L(k : \bar{p}, \bar{C}_p) = 0\right\}. \quad (13.9)$$

By choosing an appropriate significance level for the confidence limits of p and C_p (α_p and α_{C_p}, respectively), we can then determine the minimum confidence level of the

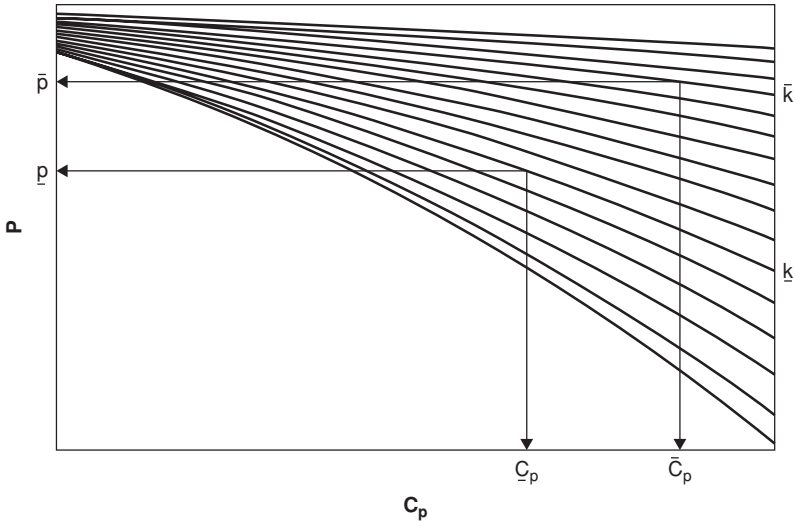


Figure 13.3 Graphical representation of equations (13.8) and (13.9).

resulting confidence limits for k by invoking the Bonferroni inequality. As commented by Kotz and Johnson,[10] such Bonferroni inequalities are usually overly conservative. This conservatism originates from the large sampling variability associated with p. For a stable, well-established process we could assume that p is known. This leads us to propose the following modifications to equations (13.8) and (13.9):

$$\underline{k} = \{k : L(k : p, \underline{C}_p) = 0\}, \tag{13.10}$$

$$\bar{k} = \{k : L(k : p, \bar{C}_p) = 0\}. \tag{13.11}$$

This assumption is not unrealistic, because the yields for most established industrial processes have been monitored closely. Graphically, the confidence limits for k can be obtained by moving along the contour of constant p as C_p varies from $\overline{C}_p$ to $\underline{C}_p$. This is shown graphically in Figure 13.4.

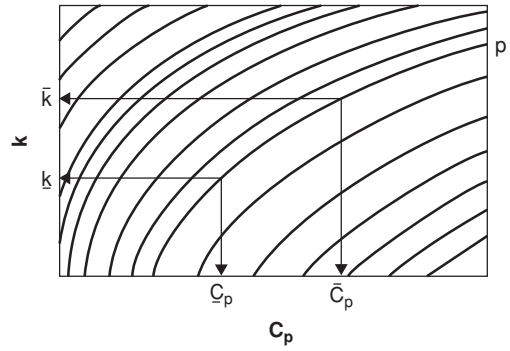


Figure 13.4 Graphical representation of equations (13.10) and (13.11).

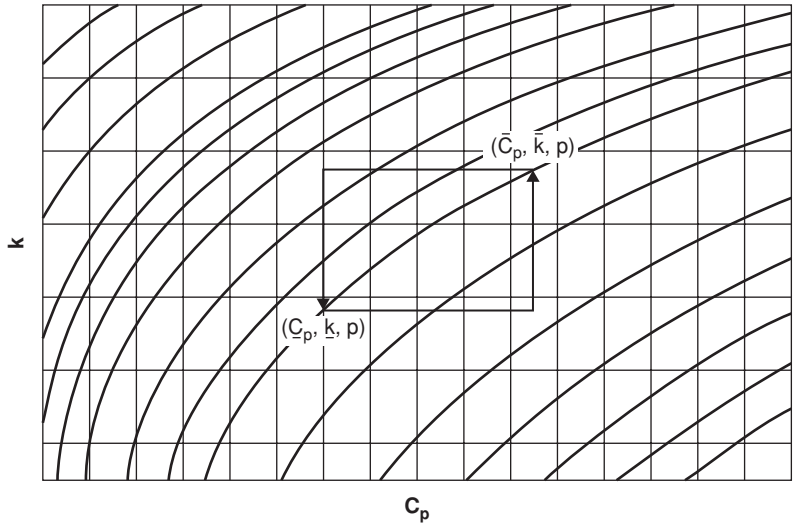


Figure 13.5 Tolerance box for computing upper and lower confidence limits of k.

Projecting horizontally across at the points from $\underline{C}_p$ to $\overline{C}_p$ intersecting the contour of $\underline{p}$, we will arrive at the estimated values of k and $\bar{k}$. From Figure 13.4 the estimated $\underline{k}$ then corresponds to $\underline{C}_p$ and $\overline{k}$ corresponds to $\overline{C}_p$. As shown in Figure 13.5, a rectangular box can also be drawn with $(\overline{C}_p, p, \overline{k})$ and $(\underline{C}_p, p, \underline{k})$ at the two opposite extreme corners. Hence the solution to equations (13.10) and (13.11) would simply imply locating the two corners of the postulated box.

13.4 CONFIDENCE LIMITS FOR C_{pk}

The index

$$C_{pk} = \frac{\min(USL - \mu, \mu - LSL)}{3\sigma} \tag{13.12}$$

gives the scaled distance between the process mean (μ) and the closest specification limit. C_{pk} can also be written in terms of k as

$$C_{pk} = (1 - k)\, C_p. \tag{13.13}$$

Having established the confidence limits for k, we propose three approximate two-sided confidence limits for C_{pk} by assuming that:

(a) variability in C_p dominates, while that of k is not significant;
(b) variability in k dominates, while that of C_p is not significant;
(c) variability in both k and C_p is significant.

13.4.1 Variability in C_p dominates while that of k is not significant

Since C_p pertains to the process spread and k to the shift in process mean, case (a) represents the practical scenario of a long-run and mature production process. In this context one has a wealth of historical data available to predict or estimate the shift in process mean accurately. With well-designed statistical experiments and process control it is possible to control the process mean at the desired target. It is the inherent variability in the process spread that is hard to eliminate; thus it dominates in the total variation of C_{pk}. Therefore it is appropriate to assume case (a) when one samples from a long-run and mature production with process mean well under control. From equation (13.13) the corresponding confidence limits for C_{pk} are given by

$$\left[\underline{C}_{pk1}, \overline{C}_{pk1}\right] = \left[(1-k)\,\underline{C}_p, (1-k)\,\overline{C}_p\right]. \tag{13.14}$$

For 95 % confidence limits for C_{pk1} we set the significant level of C_p at $\alpha_{1,C_p} = 0.05$.

13.4.2 Variability in k dominates while that of C_p is not significant

Case (b) reflects the situation where one samples from a short-run production process during start-up or pilot run. The limited amount of data available for sampling permits only a snapshot of the process at the instant when sampling is conducted. Since the typical adjustability of the process mean is higher than that of the inherent process spread, this instant snapshot is more likely to give a better and more accurate picture of the process spread than of the process mean in the short run. Variability in the process mean or index k will then be of major concern in this case when assessing the sampling variability in the index C_{pk}. The second type of confidence limit for C_{pk} is given by

$$\left[\underline{C}_{pk2}, \overline{C}_{pk2}\right] = \left[(1-\overline{k})\underline{C}_p, (1-\underline{k})C_p\right]. \tag{13.15}$$

For 95 % confidence limits for C_{pk2} we set the significance level of k at $\alpha_{2,k} = 0.05$.

13.4.3 Variability in both k and Cp is significant

For case (c) we combine the confidence limit statements of k and C_p using the Bonferroni inequality to give conservative confidence limits for the actual joint confidence level of $C_{pk.}$ This is denoted by $[\underline{C}_{pk3}, \overline{C}_{pk3}]$, where

$$\left[\underline{C}_{pk3}, \overline{C}_{pk3}\right] = [(1-\overline{k})C_p, (1-\underline{k})\overline{C}_p]. \tag{13.16}$$

The significance levels of C_p and k are set at $\alpha_{3,C_p} = 0.025$ and $\alpha_{3,k} = 0.025$ respectively, so that through the Bonferroni inequality the joint confidence level of C_{pk3} will be at least 95 %, since $(1-\alpha_{C_{pk3}}) \geq 1-\alpha_{3,C_p}-\alpha_{3,k} = 0.95$. Though this may represent the actual sampling scenario in most production processes where variability in C_p and k coexists to contribute to the total variation in C_{pk}, the Bonferroni limits are conservative.

13.5 A SIMULATION STUDY

To compare the performance of the proposed approximate confidence limits, a series of simulations was undertaken. Simulations each consisting of 10 000 random samples of size $n = 50$ were conducted to investigate the coverage probability and average width of the confidence limits. The values $USL = 40.0$, $LSL = 10.0$ and $\underline{m} = 25.0$ were used in all simulations. In each run, $\overline{x}$ and s were obtained from the 50 values generated from a normal process; $\hat{k}$, $\hat{C}_p$ and $\hat{C}_{pk}$ were calculated from $\overline{x}$ and s. A 95 % confidence interval for $\hat{C}_{pk}$ was constructed using each of the three methods discussed in Section 13.4. Each run was then replicated 10 000 times. The coverage probabilities could then be compared with the expected value determined by the significance level for C_p and k.

The results obtained are given in Tables 13.1–13.4 for $C_p = 1.0, 1.5, 2.0$ and 2.5, respectively. The value of μ is varied in each simulation run (column), so that k is set at 0.01, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3, 0.5 and 0.7. The significance level for C_p and k has been set in such a way that the expected coverage probability for C_{pk1}, C_{pk2} and C_{pk3} is 0.95. All simulations were run on a DEC 4000 with random number generation and nonlinear optimization accomplished using IMSL subroutines.

As expected, the coverage probability for case (c) is always greater than the expected value of 0.95 for all values of C_p, n and k. Its average confidence limit width is always greater than those of cases (a) and (b). This conservative property is typical of confidence limits constructed using the Bonferroni inequality. However, our main interest in this study is to examine the performance of C_{pk1}, C_{pk2} and C_{pk3} for different values of k and to determine the appropriate type of confidence limit to be used over different ranges of k.

Table 13.1 Coverage probability and average confidence limit width from simulation at 95 % confidence level ($C_p = 1.0$, $n = 50$).

	$k = 0.01$	$k = 0.03$	$k = 0.05$	$k = 0.07$	$k = 0.1$	$k = 0.2$	$k = 0.3$	$k = 0.5$	$k = 0.7$
$\underline{C}_{pk1}$	0.79446	0.77841	0.76236	0.74361	0.72223	0.64199	0.56174	0.40124	0.24075
$\overline{C}_{pk1}$	1.18546	1.16121	1.13727	1.11333	1.07741	0.95770	0.83799	0.59856	0.35914
Width	0.39100	0.38280	0.37941	0.36702	0.35518	0.31571	0.27625	0.19732	0.11839
Cov. prob.	0.9422	0.9329	0.9279	0.9166	0.9077	0.9075	0.8877	0.8408	0.7009
$\underline{C}_{pk2}$	0.77446	0.77129	0.76518	0.7565	0.7397	0.66632	0.58443	0.41766	0.25060
$\overline{C}_{pk2}$	1.00000	1.00000	1.00000	0.99999	0.9999	0.99999	1.0000	0.62455	0.37391
Width	0.22554	0.22871	0.23482	0.24349	0.26029	0.33367	0.41557	0.20689	0.12331
Cov. prob.	0.6123	0.6220	0.6646	0.6997	0.7642	0.9147	0.9400	0.8483	0.7053
$\underline{C}_{pk3}$	0.58637	0.58397	0.57935	0.57279	0.56009	0.50455	0.44255	0.31626	0.18976
$\overline{C}_{pk3}$	1.22735	1.22735	1.22735	1.22735	1.22735	1.22735	1.22735	0.79377	0.47445
Width	0.64098	0.64338	0.64800	0.65456	0.66726	0.72280	0.7848	0.47751	0.28469
Cov. prob.	0.9816	0.9838	0.9857	0.9901	0.9944	0.9999	1.0000	0.9995	0.9800

Table 13.2 Coverage probability and average confidence limit width from simulation at 95 % confidence level ($C_p = 1.5$, $n = 50$).

	$k = 0.01$	$k = 0.03$	$k = 0.05$	$k = 0.07$	$k = 0.1$	$k = 0.2$	$k = 0.3$	$k = 0.5$	$k = 0.7$
$\underline{C}_{pk1}$	1.19169	1.16761	1.14354	1.11946	1.08335	0.96298	0.84261	0.60186	0.36112
$\overline{C}_{pk1}$	1.77773	1.74182	1.70590	1.66999	1.61612	1.43655	1.25698	0.89784	0.53871
Width	0.58604	0.57421	0.56236	0.55053	0.53277	0.47357	0.41437	0.29598	0.17759
Cov. prob.	0.9436	0.9401	0.9331	0.9238	0.9243	0.9232	0.9161	0.8926	0.8179
$\underline{C}_{pk2}$	1.21004	1.20026	1.18328	1.1621	1.12676	1.00239	0.87710	0.62560	0.37590
$\overline{C}_{pk2}$	1.50000	1.50000	1.50000	1.50000	1.50000	1.50000	1.31208	0.92460	0.56076
Width	0.28996	0.29974	0.31672	0.33790	0.37324	0.49761	0.43498	0.29810	0.18486
Cov. prob.	0.5563	0.5900	0.6433	0.6948	0.7816	0.9368	0.9292	0.8971	0.8231
$\underline{C}_{pk3}$	0.91627	0.90887	0.89601	0.87998	0.85322	0.75903	0.66416	0.4744	0.28464
$\overline{C}_{pk3}$	1.84103	1.84103	1.84103	1.84103	1.84102	1.84103	1.67355	1.84102	0.71142
Width	0.92476	0.93216	0.94502	0.96105	0.98780	1.08200	1.00939	1.36662	0.42678
Cov. prob.	0.9772	0.9834	0.9842	0.9920	0.9967	0.9998	0.9997	1.0000	0.9973

In practice the confidence limits given by C_{pk3} may be too conservative. An alternative is to propose less conservative confidence limits which are easy to construct and yet have coverage probability close to the desired value. Though not included in this chapter, we have tables giving simulation results consisting of k- and C_p-values at finer resolution with sample size $n = 100$. These results are comparable and consistent with those reported here. It is important to note from Tables 13.1–13.4 that the coverage probability of C_{pk1} is close to the desired level for k ranging from 0.01 to 0.1.

Table 13.3 Coverage probability and average confidence limit width from simulation at 95 % confidence level ($C_p = 2.0$, $n = 50$).

	$k = 0.01$	$k = 0.03$	$k = 0.05$	$k = 0.07$	$k = 0.1$	$k = 0.2$	$k = 0.3$	$k = 0.5$	$k = 0.7$
$\underline{C}_{pk1}$	1.58892	1.55682	1.52472	1.49262	1.44447	1.28397	1.12348	0.80248	0.48149
$\overline{C}_{pk1}$	2.37031	2.32242	2.27454	2.22665	2.15483	1.91540	1.67598	1.19713	0.71828
Width	0.78139	0.76560	0.74982	0.73403	0.71036	0.63143	0.55250	0.39465	0.23679
Cov. prob.	0.9445	0.9330	0.9317	0.9316	0.9331	0.9291	0.9279	0.9175	0.8760
$\underline{C}_{pk2}$	1.63605	1.61573	1.58597	1.55345	1.50357	1.33653	1.16947	0.83533	0.50120
$\overline{C}_{pk2}$	2.0000	2.0000	2.0000	2.0000	2.0000	2.00000	1.74479	1.24613	0.74768
Width	0.36395	0.38427	0.41403	0.44655	0.49643	0.66347	0.57532	0.41080	0.24648
Cov. prob.	0.5198	0.5784	0.6338	0.7028	0.7899	0.9413	0.9381	0.9227	0.8839
$\underline{C}_{pk3}$	1.23886	1.22347	1.20094	1.17631	1.13854	1.01206	0.88555	0.63254	0.37952
$\overline{C}_{pk3}$	2.45470	2.45470	2.45470	2.45470	2.45470	2.4547	2.21473	1.58093	0.94856
Width	1.21584	1.23123	1.25376	1.27839	1.31616	1.44264	1.32918	0.94839	0.56904
Cov. prob.	0.9772	0.9796	0.9879	0.9929	0.9954	0.9999	0.9998	0.9996	0.9997

Table 13.4 Coverage probability and average confidence limit width from simulation at 95 % confidence level ($C_p = 2.5$, $n = 50$).

	$k = 0.01$	$k = 0.03$	$k = 0.05$	$k = 0.07$	$k = 0.1$	$k = 0.2$	$k = 0.3$	$k = 0.5$	$k = 0.7$
$\underline{C}_{pk1}$	1.98614	1.94602	1.90590	1.86577	1.80559	1.60496	1.40434	1.0031	0.60186
$\overline{C}_{pk1}$	2.96289	2.90303	2.84317	2.78332	2.69353	2.39425	2.09497	1.49641	0.89784
Width	0.97675	0.95701	0.93727	0.91755	0.88794	0.78929	0.69063	0.49331	0.29598
Cov. prob.	0.9450	0.9365	0.9349	0.9342	0.9323	0.9314	0.9278	0.9214	0.8970
$\underline{C}_{pk2}$	2.05724	2.02449	1.98379	1.94214	1.87950	1.67067	1.46183	1.04417	0.62650
$\overline{C}_{pk2}$	2.50000	2.50000	2.50000	2.50000	2.50000	2.49999	2.18074	1.55767	0.93460
Width	0.44276	0.47551	0.51621	0.55786	0.62050	0.82932	0.71891	0.51350	0.30810
Cov. prob.	0.5157	0.5652	0.6390	0.7030	0.7925	0.9428	0.9393	0.9316	0.9052
$\underline{C}_{pk3}$	1.5578	1.53300	1.50218	1.47064	1.42321	1.26507	1.10694	0.79067	0.47440
$\overline{C}_{pk3}$	3.06837	3.06837	3.06837	3.06837	3.06837	3.06837	2.76674	1.97617	1.18570
Width	1.51057	1.53537	1.56619	1.59773	1.64516	1.8033	1.65980	1.1855	0.71130
Cov. prob.	0.9728	0.9797	0.9882	0.9920	0.9958	0.9996	1.00000	0.9996	0.9997

This coverage probability approaches 0.95 as the value of C_p increases. The average width of the C_{pk1} confidence interval decreases as k increases. However, it increases as C_p increases.

For k ranging from 0.2 to 0.5, C_{pk2} outperforms C_{pk1} and C_{pk3} in that the coverage probability values are closest to the targeted level. Similar to the trend exhibited by C_{pk1}, the coverage probability of C_{pk2} approaches 0.95 as C_p increases. The average confidence limit width decreases as k increases but increases as C_p increases.

By examining the performance of the three different types of confidence limits for different k-values, we can determine the 'breakpoint' where the contribution of k to the total variation becomes dominant in C_p and can be assumed to be the sole C_{pk} total variation contributor. This occurs for k ranging from 0.2 to 0.5. The 'breakpoint' is consistently located at $k = 0.2$ even when C_p changes. A similar observation is reported for simulation at $n = 100$. Hence the 'breakpoint' is robust to the values of C_p and n.

Table 14.5 serves to determine the appropriate type of confidence limit for different ranges of k. For $k > 0.5$ we should not concern ourselves with obtaining the confidence limits for C_{pk}, since it is highly questionable whether the process itself is in statistical control. Such a large shift in the mean of the process is likely to have been detected by the control chart.

Table 13.5 Practitioner's guide to appropriate confidence limits.

	$k < 0.1$	$k = 0.1$–0.2	$k = 0.2$–0.5	$k > 0.5$
$\underline{C}_{pk}$	$\underline{C}_{pk1}$	$\underline{C}_{pk1}$ or $\underline{C}_{pk2}$	$\underline{C}_{pk2}$	Adjust process average?
$\overline{C}_{pk}$	$\overline{C}_{pk1}$	$\overline{C}_{pk1}$ or $\overline{C}_{pk2}$	$\overline{C}_{pk2}$	Adjust process average?

13.6 ILLUSTRATIVE EXAMPLES

We use the following examples to demonstrate the usage of AM confidence limits for C_{pk} under two different scenarios.

13.6.1 Example 1

A company has a well-established process for supplying a certain product for its customer. To ensure low incoming defects, the customer specifies a minimum C_{pk}-value of 1.8, with a lower confidence limit of 1.5 to cater for sampling variability. A process capability study conducted on the process reveals the following information: $n = 100, USL = 30, LSL = 12, x = 1.27$ and $s = 1.5$. Assume that the desired confidence level for all calculations is 95 % ($\alpha = 0.05$).

Substituting these data in equations (13.1)–(13.4) will give

$$\hat{C}_p = \frac{30 - 12}{6 \times 1.5} = 2.0,$$

$$m = \tfrac{1}{2}(30 + 12) = 21,$$

$$\hat{k} = \frac{|21 - 21.27|}{(30 - 12)/2} = 0.03,$$

$$\hat{C}_{pk} = (1 - 0.03)\,2.0 = 1.94.$$

Since $\hat{k} < 0.1$, Table 13.5 recommends $\underline{C}_{pk1}$as the lower AM confidence limit for C_{pk}. The k-value is reasonable in this case for a well-established and optimized process. From equation (13.6), $\underline{C}_p$ is

$$\underline{C}_p = \frac{\chi^2_{100-1,0.05/2}}{\sqrt{100 - 1}} \times 2.0 = 1.684.$$

From equation (13.14), C_{pk1} is

$$\underline{C}_{pk1} = (1 - 0.03)\,1.684 = 1.633.$$

The company can thus safely claim to the customer that this process meets the minimum C_{pk} value and lower confidence limit value.

13.6.2 Example 2

Another process is a new one acquired by a supplier. It is desired to qualify the process by running a test production. The customer specifies a minimum C_{pk} value of 1.3, with a lower confidence limit of 1.0. Since process is new, it is yet to be optimized and controlled for production. A process capability study conducted for the process records the following data: $n = 50$, $USL = 20.8$, $LSL = 10$, $x = 17.2$ and $s = 1.2$.

Substituting these data in (13.1)–(13.4) gives

$$\hat{C}_p = \frac{20.8 - 10}{6 \times 1.2} = 1.5,$$

$$m = \tfrac{1}{2}(20.8 + 10) = 15.4,$$

$$\hat{k} = \frac{|15.4 - 17.2|}{(20.8 - 10)/2} = 0.3$$

$$\hat{C}_{pk} = (1 - 0.3)\,1.5 = 1.05.$$

Since $\hat{k}$ falls between 0.2 and 0.5, Table 13.5 recommends $\underline{C}_{pk2}$ as the lower AM confidence limit for C_{pk}.

The expression for $\underline{C}_{pk2}$ in (13.15) requires the determination of $\bar{k}$. From Figure 13.4 we can determine $\overline{k}$ from p and $\overline{C}_p$, which are easily calculated from (13.5) and (13.6) respectively:

$$p = \Phi[-3(1 + 0.3)\,1.5] + \Phi[-3(1 - 0.3)\,1.5] = 8.164 \times 10^{-4},$$

$$\bar{C}_p = \frac{\chi^2_{50} - 1, 1 - 0.05/2}{\sqrt{50 - 1}} \times 1.5 = 1.796.$$

Then $\overline{k}$ can be obtained either by direct estimation from Figure 13.4 as 0.40 or by solving equation (13.11) as below using some numerical solvers:

$$\overline{k} = \left\{k : L\left(k : p = 8.164 \times 10^{-4}, \overline{C}_p = 1.796\right) = 0\right\} = 0.415.$$

From equation (13.15), using the exact k-value, $\underline{C}_{pk2}$ is determined as

$$\underline{C}_{pk2} = (1 - 0.415)\,1.5 = 0.878.$$

Clearly, we have little faith in the new process meeting the customer's quality requirement.

We can also proceed to calculate the confidence limits for the fraction nonconforming for this particular example. It is probable that the customer will be most interested in the value $\bar{p}$. From Figure 13.2 we can locate $\overline{p}$ from $\overline{k}$ and C_p (recall that $\underline{C}_{pk2} = (1 - \overline{k})C_p$), which were determined earlier as 0.415 and 1.5, respectively. Then $\overline{p}$ can be determined either by direct estimation from Figure 13.2 as approximately 4.0×10^{-3} or by calculating it from (13.5) using $\overline{k}$ and C_p as

$$\bar{p} = \Phi[-3(1 + 0.415)\,1.5] + \Phi[-3(1 - 0.415)\,1.5] = 4.24 \times 10^{-3}.$$

13.7 COMPARISON WITH BOOTSTRAP CONFIDENCE LIMITS

We have arrived at some approximate confidence limits that are computationally easy to obtain and provide adequate protection in terms of their coverage probability. In this section we compare the proposed approximate method (AM) confidence limits with three nonparametric bootstrap confidence limits for C_{pk}: the standard bootstrap (SB),

Table 13.6 Five values of parameters used in simulation study.

μ	σ	k	C_p	C_{pk}
50	2	0.0476	1.667	1.565
52	2	0.1429	1.500	0.971
52	3	0.1429	1.000	0.825
50	3.7	0.0476	0.901	0.913
52	3.7	0.1429	0.811	0.735

the percentile bootstrap (PB) and the biased-corrected percentile bootstrap (BCPB) reported by Franklin and Wasserman.[9]

The performance comparison involves a series of simulations. The values $USL = 61$, $LSL = 40$ and target $\mu = 50.5$ were used for all simulations. The five defined parameter values used in the simulation study are given in Table 13.6. These values were chosen to represent processes that vary from 'vary capable' to 'not capable'. To calculate the bootstrap confidence limits for each combination, a sample of size $n = 30$, 50 or 75 was drawn and 1000 bootstrap resamples (each of size n) were drawn from that single sample. A 90 % bootstrap lower confidence limit for C_{pk} was constructed by each of the SB, PB and BCBP methods. This single simulation was then replicated 1000 times. The frequency for each of the 90 % bootstrap confidence intervals containing the true C_{pk}-value was recorded. An average length of the bootstrap confidence limits was also calculated.

To derive the corresponding proposed AM confidence limits, we used the same parameters and number of trials (1000) as those in bootstrap for our simulations. The simulation was conducted in a similar fashion to that described in the previous section. Since all the k-values in this simulation are less than 0.01, we used $[\overline{C}_{pk1}, \underline{C}_{pk2}]$ as our AM confidence limits for all six simulation runs. To achieve a target coverage probability of 0.90, α_{1,C_p} was set at 0.10. The results of the performance comparison are given in Table 13.7.

The SB method gives coverage probabilities consistently near the expected value of 0.90. In contrast, the AM, PB and BCBP limits have coverage probabilities lower than 0.90. All three limits tend to increase slowly towards 0.90 as n increases. In general, the AM gives larger coverage probabilities than the PB and BCBP, except for the last two simulations. SB is the only one of the three bootstrap confidence intervals that consistently gives 0.90 coverage. The superiority of AM comes in when we examine the average width of the limits. As expected from the theory, the average width decreases as n increases for all limits. The AM intervals are consistently the narrowest, followed by BCBP, PB, and SB. There is a difference of approximately 10 % between the AM average width and the narrowest average width of the three bootstrap intervals.

The worst of the four methods was the PB, which has the lowest coverage probability most of the time. It has also the largest average width, which in this respect is comparable with the SB method. The AM has the smallest average width of the four methods. The difference is largest for small sample sizes and tends to reduce as $n$increases to 75. The AM also has slightly lower coverage probability than the SB method, except for low C_{pk} (<1.0) where the discrepancy increases. For high

Table 13.7 C_{pk} confidence limits: performance comparison of SB, PB, BCBP and AM.

	$n = 30$		$n = 50$		$n = 75$	
	Cov. prob.	Ave. width	Cov. prob.	Ave. width	Cov. prob.	Ave. width
$\mu = 50, \sigma = 2, k = 0.0476, C_p = 1.750, C_{pk} = 1.667, \underline{C}_{pk} = \underline{C}_{pk1}$						
SB	0.894	0.799	0.895	0.584	0.899	0.476
PB	0.841	0.785	0.856	0.579	0.875	0.474
BCBP	0.852	0.727	0.864	0.552	0.878	0.458
AM	*0.885*	*0.717*	*0.879*	*0.552*	*0.882*	*0.450*
$\mu = 52, \sigma = 2, k = 0.1429, C_p = 1.750, C_{pk} = 1.500, \underline{C}_{pk} = \underline{C}_{pk1}$						
SB	0.902	0.729	0.897	0.545	0.923	0.434
PB	0.842	0.715	0.858	0.540	0.894	0.432
BCBP	0.854	0.651	0.873	0.510	0.906	0.416
AM	*0.877*	*0.645*	*0.890*	*0.497*	0·899	*0.405*
$\mu = 50, \sigma = 3, k = 0.0476, C_p = 1.167, C_{pk} = 1.111, \underline{C}_{pk} = \underline{C}_{pk1}$						
SB	0.901	0.527	0.890	0.399	0.893	0.322
PB	0.883	0.519	0.880	0.396	0.890	0.321
BCBP	0.880	0.495	0.880	0.386	0.877	0.315
AM	*0.889*	*0.478*	*0.887*	*0.368*	*0.886*	*0.300*
$\mu = 52, \sigma = 3, k = 0.1429, C_p = 1.167, C_{pk} = 1.000, \underline{C}_{pk} = \underline{C}_{pk1}$						
SB	0.900	0.506	0.904	0.378	0.909	0.306
PB	0.825	0.496	0.866	0.375	0.884	0.306
BCBP	0.841	0.457	0.875	0.356	0.888	0.295
AM	*0.851*	*0.450*	*0.854*	*0.331*	*0.867*	*0.270*
$\mu = 50, \sigma = 3.7, k = 0.0476, C_p = 0.946, C_{pk} = 0.901, \underline{C}_{pk} = \underline{C}_{pk1}$						
SB	0.892	0.432	0.896	0.329	0.879	0.269
PB	0.875	0.425	0.887	0.327	0.880	0.268
BCBP	0.867	0.413	0.876	0.322	0.868	0.265
AM	*0.869*	*0.387*	*0.878*	*0.299*	*0.874*	*0.243*
$\mu = 52, \sigma = 3.7, k = 0.1429, C_p = 0.946, C_{pk} = 0.811, \underline{C}_{pk} = \underline{C}_{pk1}$						
SB	0.896	0.429	0.899	0.317	0.896	0.256
PB	0.849	0.423	0.862	0.315	0.867	0.256
BCBP	0.860	0.395	0.874	0.301	0.879	0.248
AM	*0.845*	*0.349*	*0.863*	*0.269*	*0.865*	*0.219*

C_{pk}-values the AM attains coverages comparable with the SB method and yet has the narrowest average widths.

13.8 CONCLUSIONS

We have proposed an approximate method to derive the confidence limits for C_{pk}. The method is easier to compute compared with other methods such as bootstrapping and non-central t integration. A graphical method was used to show the relationship between C_p, p, and k. Based on the graphs, we derived the confidence limits of k

and used the tolerance box to relate the confidence limits of C_p, p, and k. We then evaluated three approximate C_{pk} confidence limits using simulation and proposed a guide for practitioners to arrive at the 'approximate method' confidence limits. We also demonstrated their usage via illustrative examples. These AM limits were shown to have good coverage probabilities and short average confidence interval widths as compared with three bootstrap confidence limits. Their performance is robust to the sample size n and capability index C_p.

REFERENCES

1. Kane, V.E. (1986) Process capability indices. *Journal of Quality Technology*, 18, 41–52.
2. Sullivan, L.P. (1984) Reducing variability: a new approach quality. *Quality Progress*, 17(7), 15–21.
3. Montgomery, D.C. (1985) *Introduction to Statistical Quality Control*. New York: John Wiley & Sons, Inc.
4. Wang, C.W. (1990) Letters to the Editor. *Quality and Reliability Engineering International*, 6, 312–316.
5. Kushler, R. and Hurley, P. (1992) Confidence bounds for capability indices. *Journal of Quality Technology*, 24, 188–195.
6. Chou, Y.M., Owen, D.B. and Borrego, A.S.A. (1990) Lower confidence limits on process capability indices. *Journal of Quality Technology*, 22, 223–229.
7. Zhang, N.F., Stenback G.A. and Wardrop, D.M. (1990) Interval estimation of process capability index C_{pk}. *Communications in Statistics: Theory and Methods*, 19, 4455–4470.
8. Bissell, A.F. (1990) How reliable is your capability index? *Applied Statistics*, 39, 331–340.
9. Franklin, L.A. and Wasserman, G. (1991) Bootstrap confidence interval estimates of C_{pk}: an introduction. *Communication in Statistics: Simulation and Computation*, 20, 231–242.
10. Kotz, S. and Johnson, N.L. (1993) *Process Capability Indices*. London: Chapman & Hall.

14

Data Transformation for Geometrically Distributed Quality Characteristics

T. N. Goh, M. Xie and X. Y. Tang

Recently there has been increasing interest in techniques of process monitoring involving geometrically distributed quality characteristics, as many types of attribute data are neither binomial nor Poisson distributed. The geometric distribution is particularly useful for monitoring high-quality processes based on cumulative counts of conforming items. However, a geometrically distributed quantity can never be adequately approximated by a normal distribution that is typically used for setting 3σ control limits. In this chapter, some transformation techniques that are appropriate for geometrically distributed quantities are studied. Since the normal distribution assumption is used in run rules and advanced process-monitoring techniques such as the cumulative sum or exponentially weighted moving average chart, data transformation is needed. In particular, a double square root transformation which can be performed using simple spreadsheet software can be applied to transform geometrically distributed quantities with satisfactory results. Simulated and actual data are used to illustrate the advantages of this procedure.

14.1 INTRODUCTION

The geometric distribution is commonly found in process control data, and control charts based on such a distribution have received increasing attention recently. The

This chapter is based on the article by M. Xie. T. N. Goh, X. Y. Tang, 'Data transformation for geometrically distributed quality characteristics', *Quality and Reliability Engineering International*, 16(1), 2002, pp. 9–15, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

geometric-type control chart or G chart[1] is very useful when the underlying distribution is neither binomial nor Poisson and when a geometric distribution is a better description of the process behavior.

In particular, when the process is of very high quality, control charts can be developed for the cumulative count of conforming items between two nonconforming ones. The cumulative count of conforming items follows a geometric distribution. Calvin[2] first introduced the idea of tracking cumulative counts to monitor a near-zero-defect or very-high-yield process. The idea was developed into a charting technique and named by Goh.[3] Glushkovsky[1] studied the use of the G chart for process monitoring of geometrically distributed quantities, and Kaminsky *et al*.[4] also discussed the use of the geometric distribution in process monitoring. A number of examples can be found in these papers. Additional results related to the issue can be found elsewhere.[5–8]

The traditional 3σ limits for the G chart, similar to those of the p chart or c chart, have been used by Kaminsky *et al*.[4] Although 3σ control limits are convenient to use, it has been recently pointed out by Xie and Goh[9] that they are not appropriate because the geometric distribution is highly skewed. Furthermore, the normal assumption is never adequate for approximating a geometric distribution because the probability density function is always decreasing and hence cannot be bell-shaped like a normal distribution. Because of the lack of normality, supplementary run rules cannot be applied directly. Other more advanced monitoring techniques such as the standard cumulative sum (CUSUM) or exponentially weighted moving average (EWMA) chart should not be used either. The non-normality problem limits the use of the G chart. Proper transformation of the raw data needs to be studied.

A procedure is proposed by Quesenberry[5] and named the geometric Q chart. It is a form of standardized G chart. By using the transformation, the problem of detecting changes in the geometric distribution is transformed into one of monitoring a normally distributed variable so that other well-developed techniques such as supplementary run rules, CUSUM and EWMA control schemes can be used. However, the geometric Q chart control scheme assumes that the parameter is known or already estimated. Furthermore, the inverse normal is used and this is not easy to implement and interpret in practice as a plotted point has no direct meaning.

In this chapter, two other possible transformations are studied and compared with the Q-transformation. One is the simple log transformation which is a limiting case of the so-called arcsin transformation for non-normal quantities. This transformation, however, is generally not satisfactory for a geometric distribution. In line with the square root transformation for positively skewed distributions a double square root transformation is proposed, because using a single square root transformation the result is still positively skewed. Numerical results show that the normality after this transformation is usually better than with other methods. Furthermore, because the Q-transformation assumes that the model parameter is known or estimated, it is not very robust, as shown by a sensitivity analysis of the Q-transformation in the last section of this chapter.

14.2 PROBLEMS OF THREE-SIGMA LIMITS FOR THE G CHART

The idea of 3σ limits for the G chart[4] is based on the normal approximation. When the quality characteristic of interest is geometrically distributed with parameter p, the

3σ control limits are

$$UCL = \frac{1-p}{p} + \frac{3\sqrt{1-p}}{p}, \tag{14.1}$$

$$LCL = \frac{1-p}{p} - \frac{3\sqrt{1-p}}{p}. \tag{14.2}$$

This is because when Z is geometrically distributed we have that

$$F(z;p) = P(Z = z) = 1 - (1-p)^z, \qquad z = 1, 2, \ldots, \tag{14.3}$$

and the mean and variance of Z are given by

$$E(Z) = \frac{1-p}{p}, \qquad \text{Var}[Z] = \frac{1-p}{p^2}. \tag{14.4}$$

The control limits have been shown by Kaminsky *et al.*[4] to be better than the traditional c chart when the underlying distribution is geometric. However, this type of control limit has some serious problems. Xie and Goh[9] noted that for a geometric distribution the problem with the 3σ LCL is very serious as the limit always takes a negative value, which is physically meaningless. This can be seen from the fact that for the LCL in equation (14.2) to be greater than zero it is necessary that

$$\frac{1-p}{p} > 3\sqrt{\frac{1-p}{p^2}} \tag{14.5}$$

implying that $p < -8$, which is impossible as p is always positive. Note that when a G chart is used to monitor a high-quality process, p is interpreted as the process fraction nonconforming level.

To solve the problems associated with the 3σ limits of the geometric distribution, one method is to transform the individual observations so that the transformed data are more closely modeled by the normal distribution. It is then possible to proceed with the usual charting and process monitoring using the transformed data.

It should be noted that most authors using geometrically distributed quantities in process monitoring make use of the probability control limits which are more appropriate.[1,3] A drawback of this approach is that the control limits will no longer be symmetrical about the centerline. Furthermore, the advantage of process monitoring using transformed data is that run rules and other process-monitoring techniques such as EWMA and CUSUM can be applied directly as most of the procedures assume the normal distribution.

14.3 SOME POSSIBLE TRANSFORMATIONS

There are a few standard transformations, such as the square root, inverse, arcsine, inverse parabolic and square root transformations, that have been used for non-normal data.[10–13] For the geometric distribution, Quesenberry's Q transformation and the arcsine transformation have been used before. A simple, but accurate, double square root transformation is proposed in this section following a brief discussion of some other possibilities.

14.3.1 Quesenberry's Q transformation

Recently, Quesenberry[5] proposed a general method which utilizes the probability integral transformation to transform geometrically distributed data. Using Φ^{-1} to denote the inverse function of the standard normal distribution, define

$$Q_i = -\Phi^{-1}(u_i), \tag{14.6}$$

where

$$u_i = F(x_i; p) = 1 - (1-p)^{x_i}. \tag{14.7}$$

For $i = 1, 2, \ldots, Q_i$ will approximately follow the standard normal distribution, and the accuracy improves as p_0 approaches zero.

Theoretically, the Q transformation should serve the purpose well as it is based on the exact inverse normal transformation. A practical problem is that the model parameter has to be assumed known or estimated and the result could be sensitive to erroneous estimates. A sensitivity study of this issue will be carried out in a later section to further explore this issue. It is not practical to use the inverse normal transformation if it is not already built in. Furthermore, it is difficult for engineers and other chart users to interpret the results.

14.3.2 The log transformation

The a geometric distribution is a special case of the negative binomial distribution. For a negative binomial variable v with mean m and exponent k,

$$\begin{aligned} p(v) &= \frac{\Gamma(v+k)}{v!\Gamma(k)} \left(\frac{m}{m+k}\right)^v \left(1+\frac{m}{k}\right)^{-k} \\ &= \frac{(v+k-1)!}{v!\,(k-1)!} \left(\frac{m}{m+k}\right)^v \left(\frac{k}{m+k}\right)^k \end{aligned} \tag{14.8}$$

for $v = 0, 1, 2, \ldots$. Anscombe[14] showed that the corresponding transformation, which is based on a kind of inverse hyperbolic function,

$$y = \sinh^{-1}\sqrt{\frac{v+c}{k-2c}}, \tag{14.9}$$

can be used to transform a negative binomial distribution to a normal one. For the geometric distribution, as we have $k = 1$ and $v = x - 1$, the transformation becomes

$$y = \sinh^{-1}\sqrt{\frac{v+c}{1-2c}}. \tag{14.10}$$

Anscombe[14] also showed that a simpler transformation, known to have an optimum property for large m and $k \geq 1$, is

$$y = \ln\left(v + \tfrac{1}{2}k\right). \tag{14.11}$$

For the geometric distribution the transformation becomes

$$y = \ln\left(x - \tfrac{1}{2}\right).$$

In fact, when x is large, which is, for example, the case when a high-quality process is monitored based on cumulative counts of conforming items between two nonconforming ones, both the inverse hyperbolic and the log transformation are equivalent to the simple logarithmic transformation. This is because

$$\begin{aligned}\sinh^{-1} x &= \ln\left(x + \sqrt{x^2+1}\right) \\ &\approx \ln(x+x) = \ln 2 + \ln x. \end{aligned} \tag{14.12}$$

As will be seen later, the log transformation, although it can be used, is not very good for geometrically distributed quantities.

14.3.3 A double square root transformation

The method that we propose to transform a geometrically distributed quantity into a normal one is through the use of the double square root transformation

$$y = x^{1/4}, \qquad x \geq 0. \tag{14.13}$$

The background behind using the double square root transformation is as follows. It is well known that the square root transformation can be used to convert positively skewed quantities to normal. However, the geometric distribution is highly skewed, and it will be observed that after the square root transformation it is still positively skewed. In fact, we should use $x^{1/r}$ for some $r > 2$. However, considering the accuracy and simplicity, the double square root transformation is preferred as it can be performed using some simple spreadsheet software or even a simple calculator.

A number of simulation studies have been carried out, the results of which show that the double square root transformation will generally perform very well. An illustration with a set of data is shown in Figures 14.1 and 14.2. Note that Figure 14.1 has the typical shape of a geometric distribution, and without transformation it will never have a bell shape as the maximum value is zero. The transformed histogram will generally be

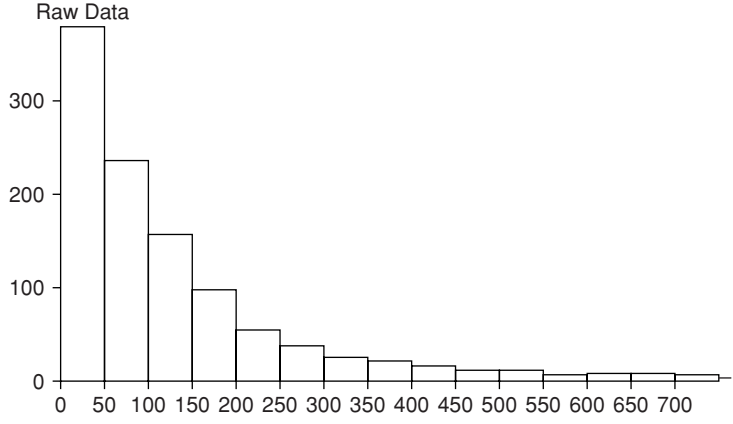


Figure 14.1 Histogram of a geometrically distributed data set.

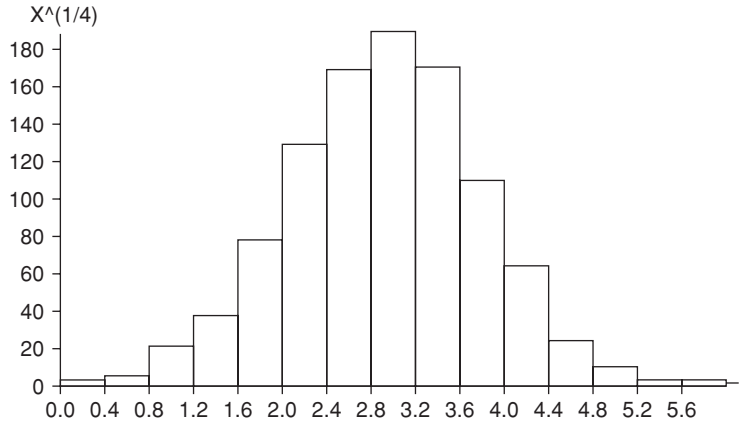


Figure 14.2 Histogram for double square root transformation of the data set in Figure 14.1.

close to a unimodal, bell-shaped curve indicating a good approximation to a normal distribution (Figure 14.2).

14.4 SOME NUMERICAL COMPARISONS

The double square root transformation performs very well in comparison with other transformations. It is easy to use. An actual data set from an electronics company that has used the transformation is shown at the end of this section. First, to illustrate the comparison with the other approaches, some data sets from Quesenberry[5] are used.

14.4.1 Example 1

A data set from Quesenberry[5] and the results of transformation are shown in Table 14.1. The Anderson–Darling test statistic and some other nonparametric statistics are used to test the normality of the distribution after the transformation. The results are shown in Table 14.2. It can be seen from the Anderson–Darling test that the normality hypothesis is not rejected for the Q transformation and double square root transformation but is rejected for the log transformation. Furthermore, the scores for the former two are very close.

To further illustrate the use of transformations when constructing a control chart in the case of process shift, we use another set of data presented by Quesenberry.[5] This is a small data set but includes a shift in process performance. The values and transformations are shown in Table 14.3.

To construct a Shewhart control chart, the first 10 transformed values are used to derive the control limits because the first 10 observations come from the same geometric distribution. The three plots are shown together in Figure 14.3.

As this is a very small example, no points are outside the control limits. However, it clearly illustrates the use of the transformation, as by the simple 'seven points below the central line' rule an alarm can be raised at point 15.

Table 14.1 A set of geometrically distributed data from Quesenberry[5] with values of transformations.

Observation	$X^{1/4}$	$\ln X$	Q	Observation	$X^{1/4}$	$\ln X$	Q
18	2.0598	2.8904	−0.2605	13	1.8988	2.5649	0.0334
1	1.0000	0.0000	1.6448	4	1.4142	1.3863	0.8946
19	2.0878	2.9444	−0.3124	64	2.8284	4.1589	−1.7801
32	2.3784	3.4657	−0.8643	11	1.8212	2.3979	0.1733
28	2.3003	3.3322	−0.7133	12	1.8612	2.4849	0.1013
8	1.6818	2.0794	0.4218	2	1.1892	0.6931	1.2959
4	1.4142	1.3863	0.8946	21	2.1407	3.0445	−0.4109
27	2.2795	3.2958	−0.6734	34	2.4147	3.5264	−0.9352
72	2.9130	4.2767	−1.9617	15	1.9680	2.7081	−0.0921
8	1.6818	2.0794	0.4218	19	2.0878	2.9444	−0.3124
41	2.5304	3.7136	−1.1646	10	1.7783	2.3026	0.2500
75	2.9428	4.3175	−2.0267	6	1.5651	1.7918	0.6282
29	2.3206	3.3673	−0.7523	7	1.6266	1.9459	0.5196
4	1.4142	1.3863	0.8946	107	3.2162	4.6728	−2.6408
11	1.8212	2.3979	0.1733	29	2.3206	3.3673	−0.7523
30	2.3403	3.4012	−0.7904	4	1.4142	1.3863	0.8946
32	2.3784	3.4657	−0.8643	36	2.4495	3.5835	−1.0036
15	1.9680	2.7081	−0.0921	1	1.0000	0.0000	1.6448
29	2.3206	3.3673	−0.7523	20	2.1147	2.9957	−0.3625
10	1.7783	2.3026	0.2500	23	2.1899	3.1355	−0.5033
25	2.2361	3.2189	−0.5906	28	2.3003	3.3322	−0.7133
4	1.4142	1.3863	0.8946	24	2.2134	3.1781	−0.5475
10	1.7783	2.3026	0.2500	55	2.7233	4.0073	−1.5586
20	2.1147	2.9957	−0.3625	7	1.6266	1.9459	0.5196
19	2.0878	2.9444	−0.3124	3	1.3161	1.0986	1.0686

Table 14.2 Comparison of three transformations using data in Table 14.1.

	Skewness	Kurtosis	Anderson–Darling statistic
$X^{1/4}$	0.1	2.8	0.30
$\ln X$	−0.6	3.1	0.69
Q-transformation	−0.2	3.0	0.31

Table 14.3 Comparison of three transformations for a set of data from Quesenberry[5] for which there is a process shift.

Observation	$X^{1/4}$	$\ln X$	Q	Observation	$X^{1/4}$	$\ln X$	Q
13766	10.8318	9.5299	−0.6669	351	4.3284	5.8593	1.8185
8903	9.7137	9.0940	−0.2262	1301	6.0058	7.1705	1.1650
35001	13.6779	10.463	−1.878	469	4.6536	6.1409	1.6907
4645	8.2556	8.4434	0.3277	1677	6.3993	7.4244	1.0177
1432	6.1516	7.2664	1.1103	47	2.6183	3.8394	2.5979
5056	8.4324	8.5282	0.2614	3298	7.5781	8.1009	0.5800
29635	13.1205	10.296	−1.6292	925	5.5149	6.8292	1.3509
11084	10.2606	9.3132	0.4397	2249	6.8865	7.7180	0.8365
2075	6.7492	7.6374	0.8875	1461	6.1825	7.2865	1.0987
183	3.6780	5.2067	2.0938	1884	6.5883	7.5408	0.9473

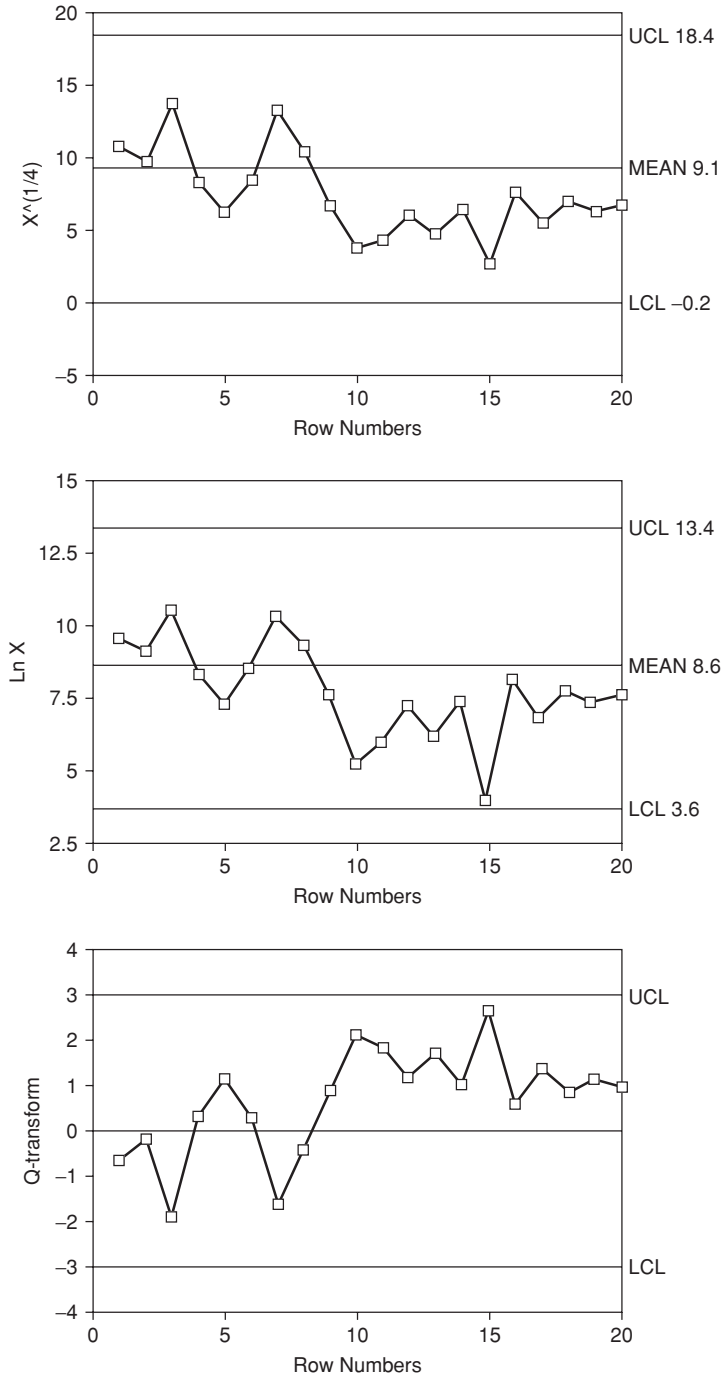


Figure 14.3 Control chart for data set in Table 14.3 based on three transformations.

Table 14.4 Comparison of three transformations.

	Skewness	Kurtosis	Anderson–Darling statistic
$X^{1/4}$	0.12*	2.27	0.90*
ln X	−0.35	2.61*	1.0
Q-transformation	−0.22	2.35	1.0

*Closest to a normal distribution on a given statistic.

14.4.2 Example 2

The double square root transformation is easily implemented in practice. In one of our projects with a local manufacturing company the problem of nonnormality was brought up. The double square root transformation has proved to be very useful. Here an actual data set of particle counts is used for illustration. The histogram has the shape of a negative exponential curve, which is the continuous counterpart of the geometric distribution. Hence the double square root transformation is thought to be appropriate.

Table 14.4 presents some normality test statistics for the data set when the three transformations are used. The double square root transformation is seen to be most suitable in terms of skewness and Anderson–Darling statistic, while for kurtosis the log transformation is preferable.

The purpose of this application was to transform the data to normal so that traditional control charts together with run rules could be used. The comparison of the original control chart and the one produced after adopting the double square root transformation is shown in Figure 14.4. It is clear that the transformed chart is more suitable in this case. It should be pointed out that a simple test of normality must be performed before using any transformation in a practical situation. Only when the normality test is accepted can run rules and techniques such as EWMA and CUSUM then be used.

14.5 SENSITIVITY ANALYSIS OF THE Q TRANSFORMATION

As discussed previously, the Q transformation should be the most appropriate transformation as it is based on the exact probability values. However, the parameter of the distribution is assumed known and it would be interesting to see how the Q transformation is affected by errors in the value of p. Recall that the Q statistic is given by equations (14.6) and (14.7). When $p = p_0$ is constant, the Q transformation can give independent and approximately standard normal statistics. Note that because of the discrete nature of the data the approximate normality depends on p_0; a smaller p_0 will give better approximations.

If an incorrect or inaccurate value of p_0 is used, the Q statistic could be totally different from what it should have been. Changes in Q statistic can be studied based on changes in p_0. This will be illustrated as follows with a numerical example.

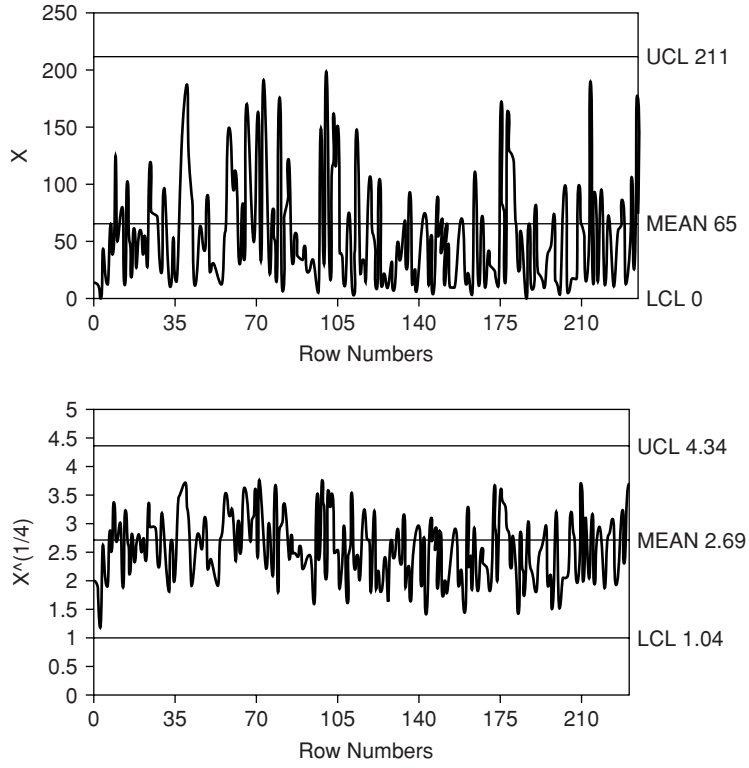


Figure 14.4 Shewhart control chart using raw data (top) and after square root transformation (bottom).

Assume that $p_0 = 0.01$ for a process. Using different p-values, the percentage change in Q-value can be calculated as

$$\frac{\Delta Q}{Q} = \frac{Q_0 - Q_1}{|Q_0|}, \tag{14.14}$$

where Q_0 represents the Q statistic using p_0 and Q_1 represents the Q statistic using p_1. Some numerical values are shown in Table 14.5.

Note that the value of the Q transformation can be very sensitive to the change in p-value. In practice, the exact value of p is not known. This means that a change in p by a factor of 50 % is not rare, especially for the G chart when it is used in high-quality process control. The same p-value should therefore not be used when the actual p has changed or when the p-values are not the same for different processes. However, it is not easy to use different p-values to transform the data in different cases. This is another limitation of the Q transformation and it makes the simple transformation more practical.

Table 14.5 Sensitivity analysis of Q when p is changed.

$\Delta p/p$	$\Delta Q/Q$	$\Delta p/p$	$\Delta Q/Q$
10 %	26 %	−10 %	−24 %
20 %	32 %	−20 %	−49 %
30 %	115 %	−30 %	−62 %
40 %	199 %	−40 %	−88 %
50 %	217 %	−50 %	−122 %
60 %	296 %	−60 %	−153 %

14.6 DISCUSSION

In this chapter the problem of nonnormality for geometrically distributed quality characteristics is discussed and some transformation techniques that can be used to transform a geometrically distributed quality characteristic to normal are studied. With the transformed data, standard statistical process control software which computes control limits based on 3σ can be used. Furthermore, procedures based on other process control techniques such as EWMA or CUSUM can be used. It seems that the Anscombe or log transformation, although it is reasonably simple and can be used for raw observations, is not appropriate for a geometric distribution.

The Q transformation of Quesenberry[5] might perform better, but assumes that the model parameter is known or estimated and is sensitive to the error of the estimated parameter. A sensitivity analysis has been conducted on the Q transformation. It seems that the Q transformation is sensitive to changes in p. The double square root transformation has been shown to be appropriate and the procedure is easily implemented.

REFERENCES

1. Glushkovsky, E.A. (1994) 'On-line' G-control chart for attribute data. *Quality and Reliability Engineering International*, 10, 217–227.
2. Calvin, T.W. (1983) Quality control techniques for 'zero-defects'. *IEEE Transactions on Components Hybrids, and Manufacturing Technology*, 6, 323–328.
3. Goh, T.N. (1987) A charting technique for low-defective production. *International Journal of Quality and Reliability Management*, 4, 18–22.
4. Kaminsky, F.C., Benneyan, J.C., Davis, R.D. and Burke, R.J. (1992) Statistical control charts based on a geometric distribution. *Journal of Quality Technology*, 24, 63–69.
5. Quesenberry, C.P. (1995) Geometric Q-charts for high quality processes. *Journal of Quality Technology*, 27, 304–315.
6. Rowlands, H. (1992) Control charts for low defect rates in the electronic manufacturing industry. *Journal of System Engineering*, 2, 143–150.
7. Bourke, P.D. (1991) Detecting a shift in fraction nonconforming using run-length control charts with 100 % inspection. *Journal of Quality Technology*, 23, 225–238.
8. Xie, M. and Goh, T.N. (1992) Some procedures for decision making in controlling high yield processes. *Quality and Reliability Engineering International*, 8, 355–360.
9. Xie, M. and Goh, T.N. (1997) The use of probability limits for process control based on geometric distribution. *International Journal of Quality and Reliability Management.*, 14, 64–73.

10. Ryan, T.P. (1989) *Statistical Methods for Quality Improvement*. New York: John Wiley & Sons, Inc.
11. Ryan, T.P. and Schwertman, N.C. (1997) Optimal limits for attributes control charts. *Journal of Quality Technology*, 29, 86–96.
12. Box, G.E.P. and Cox, D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society B*, 26, 211–243.
13. Johnson, N.L. and Kotz, S. (1969) *Discrete Distributions*. New York: John Wiley & Sons, Inc.
14. Anscombe, F.J. (1948) The transformation of Poisson, binomial, and negative binomial data. *Biometrika*, 35, 246–254.

15

Development of A Moisture Soak Model For Surface Mounted Devices

L. C. Tang and S. H. Ong

Moisture soak is a crucial step in the preconditioning test for integrated circuit packages and has a significant bearing on package quality and reliability. In this case study, we develop a moisture soak model for molded plastic leaded chip carriers with different lead counts. We first give a brief description of the problem and its motivation. Then the experimental procedure is outlined and a set of sample experimental data is presented. We then explore various models, using regression analysis, so as to establish a model that relates the moisture soak parameter to the soak conditions. A new model, different from those commonly in use in the reliability literature, has been found to provide the best fit. The equivalent soak times under harsher conditions are then estimated from the model. Finally, some comparisons are made with past results and existing industry practice.

15.1 INTRODUCTION

In electronic assembly, a high-temperature operation, *solder reflow*, is used for securing surface mounted devices (SMDs) to printed circuit boards. This operation generates a new class of quality and reliability concerns such as package delamination and cracking. The problem arises as moisture may enter the plastic molded compound through

This chapter is based on the article by L. C. Tang and S. H. Ong, 'Development of a moisture soak model', in W. R. Blischke and D. N. P. Murthy (eds), *Case Studies in Reliability and Maintenance*, 2003, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

diffusion. This moisture inside the plastic package will turn into steam, which expands rapidly when exposed to the high temperature of the vapor phase reflow infrared soldering. Under certain conditions, the pressure from this expanding moisture can cause internal delamination, internal cracks and other mechanical damage to the internal structure of the package.[1–8] Such damage can result in circuit failure, which could immediately affect the yield or could be aggravated over time, thus affecting device reliability. This has been a major quality and reliability concern for semiconductor manufacturers and users.

A proven technique in detecting various package problems is a preconditioning test as specified in JESD22-A113A (1995).[9] This crucial test is usually conducted prior to other reliability tests in order to simulate the types of processes an integrated circuit (IC) package is likely to go through before being assembled. For SMDs, the procedure includes electrical and functional tests, external visual inspection, temperature cycling, baking, moisture soak and reflow. There are various levels of moisture soak and reflow, which are determined by the classification of the package during its qualification stage. Moisture absorption and desorption tests are seldom carried out alone except for identifying the classification level of a plastic molded compound. This test is used for identifying the classification level of plastic SMDs so that they can be properly packed and stored. This will better protect the package and avoid any subsequent mechanical damage during reflow or any repair operation.

There are two basic issues in moisture absorption: the rate of absorption and the saturation level. Plastic mold compounds transport moisture primarily by diffusion. Galloway and Miles[10] tackled the absorption and desorption of moisture in plastic molded compound through the measure of moisture diffusivity and solubility as a function of time. It was established that transportation strictly by diffusion could be modeled using the standard transient diffusion equation,

$$\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2} = \frac{1}{\alpha}\frac{\partial C}{\partial t},$$

where C is the concentration level, α is the diffusivity, (x, y, z) are spatial coordinates, and t is time. Using the standard separation technique results in an expression for the local concentration, which in turn gives an analytical expression for the total weight gain as a function of time. Saturation concentration, which is a function of temperature, humidity and material, determines the maximum possible weight gain per sample volume for a particular ambient condition. This forms the basis for the functions under consideration in this case study.

Two similar international standards, EIA/JESD22-A112-A[11] and IPC-SM-786A,[12] have been developed to classify SMD packages. Both define six levels of classification. Each level corresponds to the maximum floor life that a package can be exposed to in the board assembly environment prior to solder reflow without sustaining package damage during the reflow operation. Table 15.1 shows the levels defined in JESD22-A113-A[9] for the moisture soak of the preconditioning test.

Huang *et al*.[5] found that delamination and 'popcorn' (a term commonly used to describe the failure phenomenon due to the rapid expansion of water vapour trapped in plastic packages) of plastic packages were highly dependent on reflow parameters. The dominant reflow factor affecting delamination is the total heat energy applied to the package, which is represented by a combination of the total time above 100°C

Table 15.1 Moisture sensitivity levels.

Level	Floor life	Moisture soak	
1	Unlimited at $\leq 85^{\circ}$C/85%RH	85°C/85%RH	168 hours
2	1 year at $\leq 30^{\circ}$C/60%RH	85°C/60%RH	168 hours
3	1 week at $\leq 30^{\circ}$C/60%RH	30°C/60%RH	192 hours
4	72 hours at $\leq 30^{\circ}$C/60%RH	30°C/60%RH	96 hours
5	24/48 hours at $\leq 30^{\circ}$C/60%RH	30°C/60%RH	48/72 hours
6	6 hours at $\leq 30^{\circ}$C/60%RH	30°C/60%RH	6 hours

and the temperature throughout the soak time. The higher the temperature and the longer the dwell time over 100°C, the more likely a package is to crack if the moisture level in the package is above the critical level. The study concluded that the JEDEC-recommended vapor phase and convection reflow temperature preconditioning profiles are adequate guidelines to determine the moisture sensitivity of plastic packages, unless excessively long preheat is applied. This forms the basis for using JEDEC as the benchmark in our study.

From Table 15.1, it can be seen that there are six levels of moisture soak with three different temperature–humidity settings. Therefore preconditioning testing of different products with different moisture sensitivity levels requires three chambers if they are to be done at the same time. Otherwise, we will have to complete the preconditioning test at one temperature–humidity setting before proceeding to another test with a different setting. This will greatly increase the cycle time of the test, thus reducing productivity. Since moisture soaking takes much longer than the other precondition tests, it is of great value to see if the equivalent moisture soak times at harsher temperature–humidity conditions for levels 2–6 can be obtained. Given its potential economic value in terms of lower capital investment as well as shorter time to market, an investigation into the rate of moisture absorption on a selected set of devices is conducted. The aim is to develop a model for moisture gain under various conditions, from which the equivalent moisture soak time at a harsher condition can be derived. Specifically, a suitable response related to the weight gain process will be identified and a family of response functions under pre-set conditions together with a physical model will be established. Analogous to the basic framework of accelerated testing models, the moisture soaking acceleration factor, and hence the equivalent moisture soak time, for a preconditioning test on a plastic package can be derived.

In the following section, we outline the experimental procedure and present a set of sample results. In Section 15.3 various choices for the moisture soak model are discussed and a regression analysis is performed to establish the most appropriate alternative. A set of acceleration factors is generated from the model. Finally, some discussions and conclusions are given in Section 15.4.

15.2 EXPERIMENTAL PROCEDURE AND RESULTS

The plastic leaded chip carrier (PLCC) package with B24 mold compound was selected as the primary test specimen for the case study. Three lots of 44-, 68-, and 84-lead (pin)

Table 15.2 Experimental units.

Lead count	44	68	84
Die size	120 × 120 mil	148 × 160 mil	210 × 210 mil
DAP size	250 × 250 mil	280 × 280 mil	300 × 300 mil

PLCCs were manufactured. The die and die attach pad (DAP) size were maximized so as to create a larger critical surface area and allow a higher likelihood of delamination. These are summarized in Table 15.2.

The tests were carried out in accordance with the procedures and standard set by JEDEC, EIA/JESD22-A112A[11] and National Semiconductor specifications. The flowchart for a moisture absorption test is given in Figure 15.1.

Votsch temperature–humidity chambers and a Mettler precision weighing scale were used for the moisture absorption and desorption tests. Before the actual commencement of the moisture test, calibrations for all the equipment were carried out. Since the weighing system is very important for this study, a gage repeatability and reproducibility was carried out to ensure the capability of the weighing system.

Four samples of 20 units each were prepared for the four temperature–humidity test conditions of 30°C/60% RH, 60°C/60% RH, 85°C/60% RH, and 85°C/85% RH.

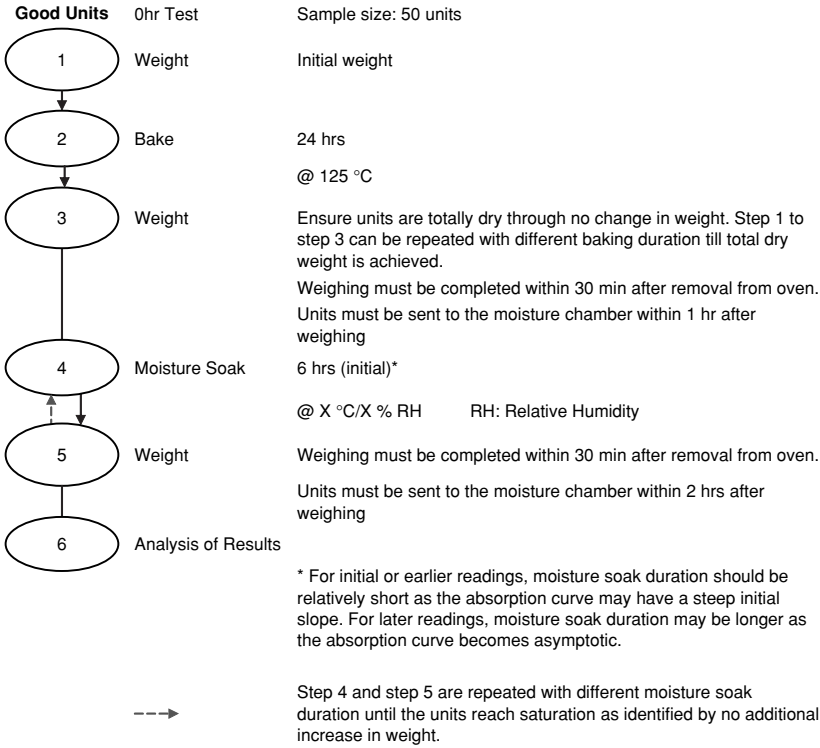

Good Units
0hr Test
Sample size: 50 units
1
Weight
Initial weight
2
Bake
24 hrs
@ 125 °C
3
Weight
Ensure units are totally dry through no change in weight. Step 1 to step 3 can be repeated with different baking duration till total dry weight is achieved.
Weighing must be completed within 30 min after removal from oven.
Units must be sent to the moisture chamber within 1 hr after weighing
4
Moisture Soak
6 hrs (initial)*
@ X °C/X % RH RH: Relative Humidity
5
Weight
Weighing must be completed within 30 min after removal from oven.
Units must be sent to the moisture chamber within 2 hrs after weighing
6
Analysis of Results
* For initial or earlier readings, moisture soak duration should be relatively short as the absorption curve may have a steep initial slope. For later readings, moisture soak duration may be longer as the absorption curve becomes asymptotic.
Step 4 and step 5 are repeated with different moisture soak duration until the units reach saturation as identified by no additional increase in weight.

Figure 15.1 Flowchart for moisture absorption test.

Table 15.3 Typical data for moisture absorption test.

84-lead PLCC Moisture Absorption Data (all weights are in grams)					
Soak Condition: 85°C /85%RH			Bake Condition: 125°C		
Duration(hr)	Weight	Weight	Avg. weight	Weight gain	% Weight gain
0	123.6302	123.6307	123.6305	0.0000	0.0000
4	123.7218	123.7229	123.7224	0.0919	0.0743
8	123.7578	123.7578	123.7578	0.1274	0.1030
24	123.8451	123.8467	123.8459	0.2155	0.1743
48	123.9153	123.9156	123.9155	0.2850	0.2305
72	123.9500	123.9511	123.9506	0.3201	0.2589
96	123.9734	123.9740	123.9737	0.3433	0.2776
120	123.9858	123.9865	123.9862	0.3557	0.2877
144	123.9968	123.9971	123.9970	0.3665	0.2964
168	124.0040	124.0045	124.0043	0.3738	0.3024

Typically, units are only required to undergo a 24-hour baking in a 125°C oven. However, due to the importance of accuracy in the weight of the units, all the units were over-baked to ensure that they have reached their absolute dry weight in a 125°C baking environment.

The moisture gain data for the four test conditions were recorded at regular intervals. The initial sampling interval used for the recording of the moisture gain data differs between different test conditions. The initial sampling interval is shorter for the harsher test condition to allow for an anticipated higher moisture absorption rate. Test units did not undergo moisture soaking till saturation, as this was not required. Units that underwent moisture soaking with conditions harsher than 30°C/60% RH only lasted 168 hours, as the moisture gain was already much more than the 30°C/60% RH test condition.

When test units were removed from chambers to be weighed, additional effort was made to ensure the consistency of every action, from the loading/unloading time to the steps taken in weighing the units (including the placement of the units in the weighing machine balance). Two sets of readings were taken for all tests and their average was used for subsequent analysis. Table 15.3 shows a typical data summary.

The data correspond to the standard moisture absorption response of a level 4 mold compound. For comparison of moisture absorption data between the different lead counts, the percentage weight gain provided a better overview of the effects of the four different test conditions on the test specimen. Figure 15.2 depicts the moisture absorption for all packages with different lead counts.

15.3 MOISTURE SOAK MODEL

15.3.1 Choosing a suitable response

The logical step in constructing the moisture soak model is to fit a family of response curves to the data from which equivalent soak times under various conditions can be

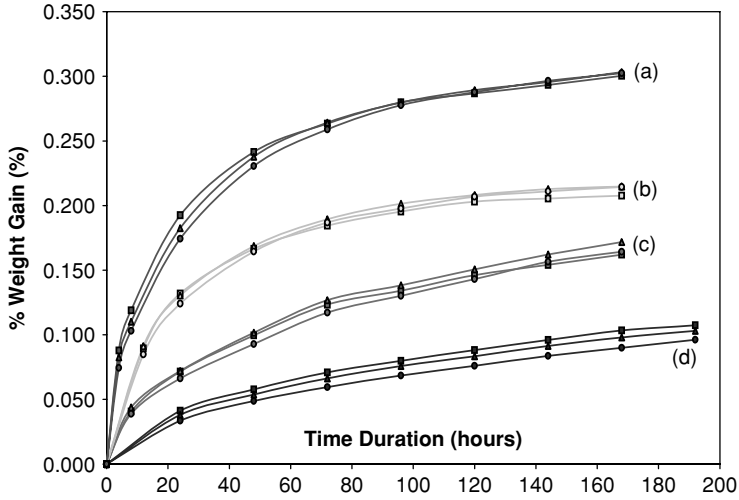


Figure 15.2 Integrated moisture absorption graph; the conditions are (a) 85/85; (b)85/60; (c) 60/60 and (d) 30/60 for three different packaging types.

estimated. There are several possible choices for the response: (a) the experimental weight gain, (b) the ratio of experimental weight gain to the maximum weight gain of the specimen, and (c) the ratio of experimental weight gain to the initial weight of the specimen. To eliminate possible dependency of the weight gain on any specific specimen, the use of a ratio, (b) or (c), is preferred. But the use of (b) requires that all specimens undergo moisture soaking till full saturation, which would not only consume too much equipment time but also introduce additional sources of experimental errors due to prolonging the experiment. As a result, (c) is adopted and is listed as '% weight gain' together with the experimental data in Table 15.3. Moreover, normalizing the weight gain by the initial weight also allows for comparison of moisture absorption data between packages of different lead counts. This will provide a better overview of the effects of the four different test conditions on the test specimens. Figure 15.2 depicts the moisture absorption as percentage weight gain for all three packages with different lead counts.

15.3.2 Choosing the family of response curve

Let W_t denote the weight gain at time t and W_0 the initial weight of a package. The reciprocal of the response chosen, W_0/W_t, is akin to a reliability function although, in practice, weight gain will not be infinite. Nevertheless, the same method for assessing goodness of fit as in probability plotting can be adopted using the common choices of reliability functions. Preliminary exploration reveals that of the exponential, Weibull, logistic, loglogistic, normal and lognormal distributions,[13] the Weibull and loglogistic distributions provide the best fit. The Weibull reliability function is given by

$$R(t) = \frac{W_0}{W_t} = \exp\left[-\left(\frac{t}{b}\right)^c\right] \Rightarrow \ln\left[\ln\left[\frac{W_t}{W_0}\right]\right] = c\ln(t) - \ln(b), \tag{15.1}$$

where b is the scale parameter and c is the shape parameter.

The loglogistic reliability function is given by

$$R(t) = \frac{W_0}{W_t} = \frac{\exp\left[-\left(\ln(t) - a\right)/b\right]}{1 + \exp\left[-\left(\ln(t) - a\right)/b\right]} \Rightarrow \ln\left[\frac{W_t - W_0}{W_0}\right] = \left(\frac{\ln(t) - a}{b}\right), \tag{15.2}$$

where a is the location parameter and b is the scale parameter. The good fit to both distributions is expected as, when $W_t - W_0$ is small, we have

$$\ln\left[\frac{W_t}{W_0}\right] \approx \frac{W_t - W_0}{W_0}.$$

The loglogistic and Weibull plots, resulting from equations (15.1) and (15.2), are given in Figures 15.3. It will be evident that both the Weibull and loglogistic provide good fits, as the plots are nearly linear. The goodness of fit to the loglogistic distribution will be further reinforced in the combined analysis presented in the following section. It is also noted that the plot for the loglogistic under different experimental conditions

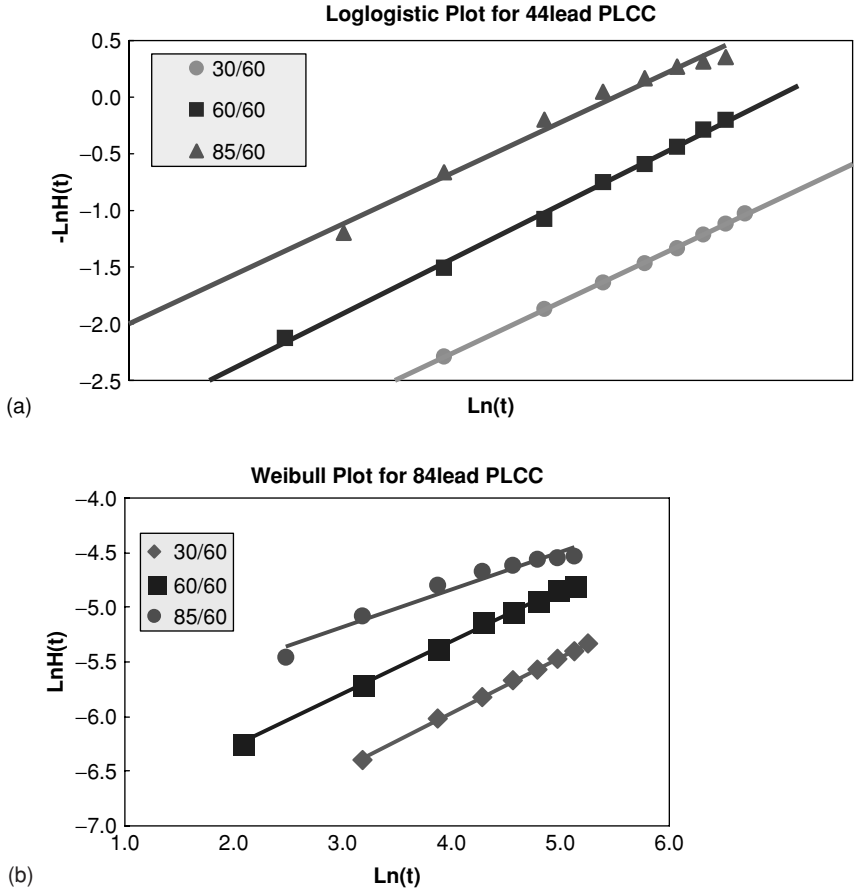


Figure 15.3 Probability plots: (a) loglogistic; (b) Weibull.

gives nearly parallel lines. This justifies the use of a common scale parameter, b, for different experimental conditions.

In the following, we shall use the loglogistic distribution for further analysis and discussion. However, it should be borne in mind that the same principle can be extended to other distributions.

15.3.3 Combined analysis with the acceleration model

In order to estimate the equivalent moisture soak time for conditions other than those tested, we need to establish an acceleration model that best fits the moisture gain data under various relative humidity (RH) and temperature (T) combinations. From the reliability engineering literature, two models are commonly used for describing the relation between 'lifetime' and the two stress variables (T and RH): the Peck model and the generalized Eyring model. These models express the 'nominal' product life as some function of the T and RH. The exact forms of these models are given on pp. 98–102 of Nelson's book.[13] In the case of moisture absorption, since the time to reach saturation is akin to the time to failure, the logarithm of the "nominal" moisture soak time is equivalent to the location parameter a. In particular, for the Peck model, we have

$$L = A\left[\exp\left(\frac{E}{kT}\right)\right]\left[RH^{-n}\right], \tag{15.3}$$

where L is the lifetime, A is a constant, E is the activation energy and k is the Boltzmann constant. (These also apply to subsequent equations.) In the current context, with α_i $(i = 0, 1, 2)$ as the coefficients in a linear model, we have

$$a = \alpha_0 + \alpha_1\frac{1}{kT} + \alpha_2\ln(RH). \tag{15.4}$$

For the generalized Eyring model, we have

$$L = \frac{A}{T}\left[\exp\left(\frac{E}{kT}\right)\right]\exp\left[RH\left(B + \frac{C}{kT}\right)\right], \tag{15.5}$$

where B and C are constant. In the current context, this becomes

$$a + \ln(T) = \alpha_0 + \alpha_1\frac{1}{kT} + \alpha_2 RH + \alpha_3\frac{RH}{T}. \tag{15.6}$$

A variant of the Eyring model that has been used by Intel (see p. 102 of Nelson[13]) is

$$L = A\left[\exp\left(\frac{E}{kT}\right)\right][\exp(-B.RH)]. \tag{15.7}$$

This gives

$$a = \alpha_0 + \alpha_1\frac{1}{kT} + \alpha_2 RH. \tag{15.8}$$

In general, one could express the location parameter as a linear function of ln(*T*), *1/T, RH,* ln*(RH), RH/T* or other similar independent variables which are variants of the above forms. This results in

$$a = f\left(\frac{1}{T}, RH, \ln(T)\ln(RH), \frac{RH}{T}\right). \tag{15.9}$$

Here we adopt a 'combined' analysis given that the loglogistic distribution function provides the best fit. From equation (15.2), we have

$$a = -b\ln\left(\frac{W_t - W_0}{W_0}\right) + \ln(t). \tag{15.10}$$

It follows that the generic form is

$$\ln\left(\frac{W_t - W_0}{W_0}\right) = \alpha_0 + \alpha_1\ln(t) + f\left(\frac{1}{kT}, RH, \ln(T), \frac{RH}{T}, \ln(RH)\right). \tag{15.11}$$

Regression runs for (15.11), where $f(\cdot)$ takes the form of (15.4), (15.6), (15.8), are conducted. In addition, both step wise regression and best-subset regression are conducted for (15.11) to identify the best linear $f(\cdot)$. In order to investigate the effect of lead counts, a dummy variable, P, representing the type of PLCC package is also considered. The results are summarized in Table 15.4, with the coefficient of determination, R^2, and the residual root mean square, s. It can be seen that the adjusted R^2 is the best from the best-subset routine and the corresponding residual root mean square is also the smallest. Moreover, the Mallow C_p statistic[14] from the best-subset run is 5, indicating that the current set of independent variables result in a residual mean square, s^2, which is approximately an unbiased estimate of the error variance.

The resulting model is

$$\ln\left(\frac{W_t - W_0}{W_0}\right) = -49.0 + 0.390\ln(t) - 0.0302.P + 5.00\frac{RH}{T} + 6.82\ln(T), \tag{15.12}$$

where t is time in hours, $P = (-1, 0, 1)$ represents the three package types, 44-lead, 68-lead, and 84-lead respectively, RH is percentage relative humidity, and T is temperature in kelvin. The p-values for all the above independent variables are less than 0.005.

Residual analysis is carried out with plots for residuals against fitted values and residuals against observation orders (Figure 15.4). The latter plot is quite random but the former shows an obvious quadratic trend. Nevertheless, given the rather limited experimental conditions and a reasonably high $R^2(= 0.977)$, to avoid over-fitting, the

Table 15.4 Summary of results from combined analysis.

Model	Independent variables	Adjusted R^2	RMS, s
Peck	*1/T,* ln*(RH),* ln*(t)*	0.975	0.08986
Generalized Eyring	ln*(T), 1/T, RH, RH/T,* ln*(t)*	0.975	0.08988
Intel	*1/T, RH,* ln*(t)*	0.975	0.08986
Stepwise	ln*(T), RH/T,* ln*(t)*	0.9753	0.0894
Best subset	ln*(T), RH/T,* ln*(t), P*	0.977	0.08625

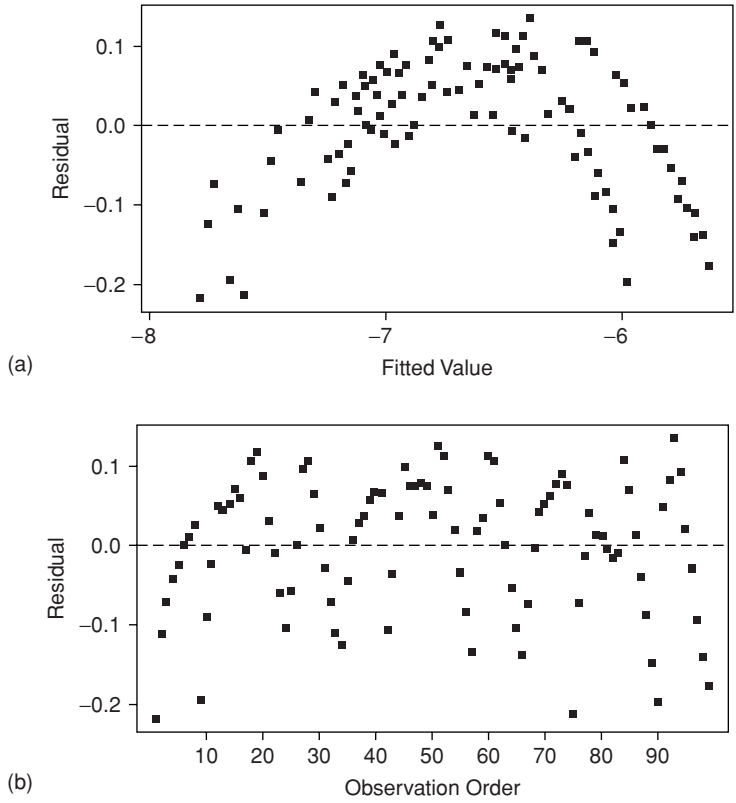


Figure 15.4 Residuals plots against (a) fitted values; (b) observation order.

current model is adequate for the experimental regime, (30–85 °C and 60–85 % RH) under consideration.

15.3.4 Calculation of equivalent moisture soak time

Combining equations (15.10) and (15.12), we can express the location parameter of the family of loglogistic distributions as

$$a = \ln(t) - 2.564 \ln\left(\frac{W_t - W_0}{W_0}\right) = 125.64 + 0.0774 \cdot P - 12.82\frac{RH}{T} - 17.49 \ln(T). \tag{15.13}$$

From (15.13), the estimate for the common scale parameter, b, for the family of loglogistic distributions is 2.564 (= 1/0.39 from (15.12); see also (15.10) and (15.2)). The location parameters and their respective standard errors under various experimental conditions are given in Table 15.5. Using the same principle as in accelerated testing, the acceleration factor between two testing conditions can be computed by taking the

Table 15.5 Predicted values of location parameter.

Package	Temperature °C	Relative humidity (%)	Location parameter, a	SE of a
1	30	60	23.1608	0.0501
1	60	60	21.7372	0.0378
1	85	60	20.6315	0.0481
1	85	85	19.7348	0.0503
0	30	60	23.0834	0.0421
0	60	60	21.6597	0.0263
0	85	60	20.554	0.0398
0	85	85	19.6573	0.0424
−1	30	60	23.0059	0.0501
−1	60	60	21.5822	0.0378
−1	85	60	20.4765	0.0481
−1	85	85	19.5799	0.0503

ratio of their respective times taken in achieving the same weight gain; which, in turn, can be expressed as the exponent of the difference between the respective location parameters. For the same package type, the acceleration factor, AF, is given by

$$AF = \exp(a_1 - a_2) = \left(\frac{T_2}{T_1}\right)^{17.49} \exp\left[-12.82\left(\frac{RH_1}{T_1} - \frac{RH_2}{T_2}\right)\right]. \tag{15.14}$$

From (15.14), the relative acceleration factor for the four testing conditions can be computed. The results are shown in Table 15.6. It should be noted that the point estimate of AF is independent of package type; however, the interval estimates are dependent on package type due to different standard errors (see the last column of Table 15.5). Given the AF, the equivalent soak time under a higher temperature–RH combination can be estimated. For example, if a level 3 package is subject to 60/60 soaking condition instead of 30/60, the soak time will be 46.3 hours instead of 192 hours as recommended by the JEDEC – a reduction by a factor of 4.15.

An approximate confidence interval for the acceleration factor could be obtained from the confidence interval for $a_1 - a_2$ using normal approximation. Specifically, from Table 15.5, the standard errors for the differences between the location parameters

Table 15.6 Acceleration factors for the current testing conditions with 95% confidence limits.

	Testing conditions in °C/%RH		
Temp(°C)/%RH combination	85/85	85/60	60/60
30/60 (levels 3,4,5,6)	30.75 (26.76,35.34)	12.54 (10.95,14.37)	4.15 (3.67,4.70)
60/60	7.41 (6.55,8.38)	3.02 (2.68,3.41)	1.00
85/60 (level 2)	2.45 (2.14,2.81)	1.00	–
85/85 (level 1)	1.00	–	–

can be computed. Then the $100(1-\gamma)\%$ confidence interval for AF is given by

$$(\underline{AF}, \overline{AF}) = \exp\left[a_1 - a_2 \pm Z\gamma/2 \cdot SE_{a_1-a_2}\right]$$
$$= \exp\left[a_1 - a_2 \pm Z\gamma/2.\sqrt{(SE_{a1})^2 + (SE_{a2})^2}\right]. \quad (15.15)$$

The results for package type 1 (44-lead) with $\gamma = 0.05$ are given in the parentheses within their respective cells in Table 15.6.

15.4 DISCUSSION

In practice, many factors need to be considered for determining the moisture-induced crack of SMD packages. Some of the more important factors are: the relative humidity and temperature of the environment, the die size, relative thickness of epoxy resin in the chip and under the chip pad, the properties of the materials, and the adhesion strength of the interfaces in the package.[15] In this study, we used PLCC packages with high pin counts. The relative humidity, temperature and soak time were the test parameters. In the following, we compare our results with past work and the prevailing industry practice.

For comparison of weight gain data, sample data from Shook *et al*.[16] for the 68-pin PLCC with the same soak condition and time scale are extracted and shown in Table 15.7. These are almost identical considering measurement error and possible differences in materials used.

Next we look at the acceleration factors given in Table 15.6. This is crucial as the equivalent moisture soak times are computed from these values. The result is well aligned with the order of stress levels. The current practice for soaking level 3–5 devices at 60°C/60% RH instead of the 30°C/60 % RH recommended by JEDEC is to use an accelerated factor of around 4.[3,4] This compares favorably with the acceleration factor of 4.15 and it also falls within the 95% confidence interval (see Table 15.6) given by our model. This, however, is not quite in agreement with Shook *et al*.[16], who suggest that testing at 60°C/60%RH will reduce the total required moisture soak time for levels 3–5 by a factor of 4.8 (our upper 97.5% confidence limit is 4.58) as compared to the time required at 30°C/60%RH. Their result is based on the empirical evidence that the diffusion coefficient at 60°C increases the kinetics of moisture ingress by a factor of 4.8 compared to that at 30°C. They then validated the result by experimenting at 30°C/60%RH and 60C/60%RH. Our experiment, however, covers a wider range of temperature–RH combinations. Our model in equations (15.12) and (15.13) also reveals that interaction between RH and T is significant. On the other

Table 15.7 Percent Weight Gain Measurements (68-pin PLCC at 30 °C/60 % RH).

	48hrs	72 hrs	96 hrs	192 hrs
Shook *et al*.	0.047	0.057	0.063	0.085
Current work	0.046	0.056	0.064	0.087

hand, it is possible that the difference in results can be due to the use of different mold compounds and some variations in the experimental process. It is widely believed that moisture absorption is dependent on the mold compound property, the amount of compound encapsulated for package, die attach epoxy, packaging construction. For different types of packaging, the moisture content of a package is dependent on the temperature and humidity. Nevertheless, from Figure 15.2, it is observed that the dependency on temperature is more pronounced. In particular, at 60%RH, the moisture absorption tends to be a linear function of temperature after an initial phase of rapid gain.

Next we compare the accelerated factor obtained for 85°C/85%RH and 30°C/60%RH. The acceleration factor of 30 is considered high compared to the value of 15 reported by Sun *et al.*,[17] who assumed a mixture of two power law models for the acceleration model without assessing the goodness of fit. In comparison with the moisture absorption graphs currently in use in industry (see Figure 15.5), our result is closer to industry practice. It has been established from the projected line that a soak time of 150 hours at 30°C/60%RH will achieve the same moisture gain as a soak time of 5 hours at 85°C/85%RH. In any case, since the floor life for level 1 is unlimited, and that for level 2 is 1 year, the industry is primarily interested in evaluating the acceleration for soaking at 60°C/60%RH for replacing the recommended 30°C/60%RH for levels 3–5. The accuracy for the acceleration factor at 85°C/85%RH is thus immaterial in practice. Moreover, it has been cautioned by Taylor *et al.*[8] that certain level 3–6 packages are not designed to tolerate the high stress soak time of 85°C/85%RH, as they fail after infraed heating when the moisture content in PQFPs is higher than 60% of the equilibrium composition.

In summary, moisture soak is a front-runner for the pre-conditioning test to assess the moisture sensitivity level during product qualification and can also be used as

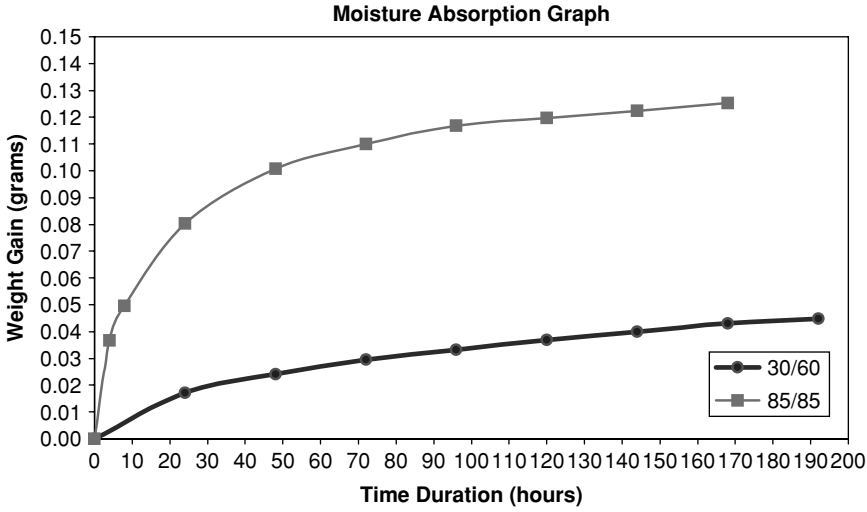


Figure 15.5 Moisture absorption graph for 30°C/60RH and 85°C/85RH temperature–humidity setting.

an in-line process monitor. To avoid having reliability and quality issues later, it is critical that moisture sensitive devices be properly classified, identified and packaged in dry bags until ready for printed circuit board assembly. Several attempts have been made to derive models, through analytic or statistical means, for predicting the soak time and humidity condition for SMDs. In this case study, we have used a statistical approach with physical reasoning to model the moisture absorption process so as to derive the equivalent moisture soak time for PLCC packages. In particular, it is found that the soak condition at 60°C/60%RH will reduce the soak time for level 3–5 (30°C/60%RH) by 3.7–4.7 times.

REFERENCES

1. Kitano, M., Nishimura, A. and Kawai, S. (1988) Analysis of the packaging cracking during reflow soldering process. *IEEE International Reliability Physics Symposium*, pp. 90–95.
2. Tay, A.A.O. and Lin, T. (1998) Moisture-induced interfacial delamination Growth in plastic packages during solder reflow. *Proceeding of the 48th Electronic Components and Technology Conference*, pp. 371–378.
3. Song, J. (1998) *Moisture Sensitivity Rating of the PLCC-68L Package*. National Semiconductor Corporation.
4. Song, J. and Walberg, R. (1998) *A Study of Standard vs. Accelerated Soak Conditions at Levels 2A and 3*. National Semiconductor Corporation.
5. Huang, Y.E., Hagen, D., Dody, G., and Burnette, T. (1998) Effect of solder reflow temperature profile on plastic package delamination. *IEEE/CPMT International Electronics Manufacturing Technology Symposium*, pp. 105–111.
6. Ahn, S.H. and Kwon, Y.S. (1995) Popcorn phenomena in a ball grid array package. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, **18B**, 491–495.
7. Tanaka, N. and Nishimura, A. (1995) Measurement of IC molding compound adhesive strength and prediction of interface delamination within package. *ASME EEP, Advances in Electronic Packaging*, 10(2):765–773.
8. Taylor, S.A., Chen, K. and Mahajan, R. (1997) Moisture migration and cracking in plastic quad flat packages. *Journal of Electronic Packaging,* **119**, 85–88.
9. JESD22-A113A (1995) *Preconditioning of Plastic Surface Mount Devices Prior to Reliability Testing*. Electronic Industrial Association.
10. Galloway, J.E. and Miles, B.M. (1997) Moisture absorption and desorption predictions for plastic ball grid array packages. *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, **20**, 274–279.
11. EIA/JESD22-A112A (1995) *Moisture-Induced Stress Sensitivity for Plastic Surface Mount Devices.* Electronic Industrial Association.
12. IPC-SM-786A (1995) *Procedures for Characterizing and Handling of Moisture/Reflow Sensitive ICs.* Institute of the Interconnecting and Packaging Electronics Circuits.
13. Nelson, W. (1990) *Accelerated Testing: Statistical Models, Test Plans and Data Analysis.* New york: John Wiley & Sons, Inc.
14. Draper, N. and Smith, H. (1981) *Applied Regression Analysis*, 2nd edition. New York: John Wiley & Sons, Inc.
15. Lee, H. and Earmme, Y.Y. (1996) A fracture mechanics analysis of the effects of material properties and geometries of components on various types of package cracks. *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, **19A**, 168–178.
16. Shook, R.L., Vaccaro, B.T. and Gerlach, D.L. (1998) Method for equivalent acceleration of JEDEC/IPC moisture sensitivity levels. Proc. 36th Int. Rel Phys. Symp., pp 214–219.
17. Sun, Y., Wong, E.C. and Ng, C.H. (1997) A study on accelerated preconditioning test. *Proceeding of the 1st Electronic Packaging Technology Conference*, pp. 98–101.

Part D

Improve Phase

16

A Glossary for Design of Experiments with Examples

H. S. Yam

Fundamental concepts in the design of engineering experiments are usually conveyed through *factorial designs* and *fractional factorial designs* which allow for the study of two or more treatment factors in reduced time and at reduced cost. This chapter provides a rudimentary outline of factorial experimental designs – comparing 2^k designs vs. 3^k designs, coded values vs. original values, standard order vs. run order, and main effects vs. interactions. It also includes terms used in the analysis of results, residual analysis and types of design. This is followed by *fractional factorial designs,* alias structure and resolution of an experiment; all of which are illustrated with examples. Finally, basic terms used in robust design such as Taguchi's *outer array of noise factors*, *inner array of control factors, repeated measures*, *replicates* and *blocking* are presented with examples.

16.1 FACTORIAL DESIGNS

Factorial designs allow for the simultaneous study of two or more treatment factors and their interaction effects. This facilitates the collection of relevant information in reduced time and at reduced cost. In the following we first present some basic terminology.

16.1.1 2^k Designs vs. 3^k Designs

In a 2^k factorial design, each factor is run at only two levels – usually the upper and lower ends of the current process window. In a 3^k factorial design, each factor is run at the upper and lowers ends, as well as the center point between them.

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

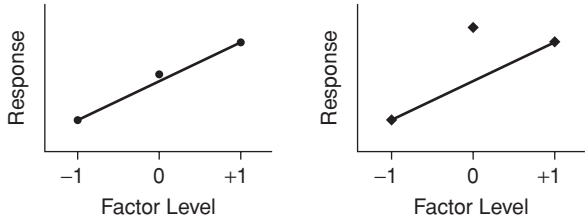


Figure 16.1 The role of 3^k factorial designs in identifying curvature.

At earlier stages of a process optimization, 2^k *factorial designs* may be sufficient as the assumption of linearity in effect is probably relevant. For mature processes, 3^k factorial designs may be more appropriate, as the assumption of linearity may no longer be valid (Figure 16.1). In situations involving both quantitative and qualitative factors, *general* 2^m3^n *factorial designs* may be appropriate.

16.1.2 Coded values vs. original values

The factor levels are often *coded* using the formula

$$x_{\text{coded}} = \frac{x - \text{Mid-point}}{\frac{1}{2}\text{Range}} = \frac{x - \frac{1}{2}(\text{Max} + \text{Min})}{\frac{1}{2}(\text{Max} - \text{Min})}.$$

Hence, the factor levels are typically set at ± 1 for the upper and lower ends, and 0 for the center point. Assignment of coded values to qualitative factors is arbitrary.

The factors' coded values are used for convenience of computation and comparison of effects between different factors. When communicating with line personnel for actual execution of the experiments, the original values should be used.

16.1.3 Standard order vs. run order

A *standard order* of runs in a 2^k factorial design has the levels for the first factor alternating for each run $(-+-+-+\ldots)$, the levels of the second alternating for each pair of runs $(--++--++\ldots)$, and so on, as illustrated in Table 16.1. The

Table 16.1 Comparison of standard order and run order.

Std order	Run order	Factor A	Factor B	Factor C
1	4	−1	−1	−1
2	6	+1	−1	−1
3	2	−1	+1	−1
4	8	+1	+1	−1
5	3	−1	−1	+1
6	7	+1	−1	+1
7	1	−1	+1	+1
8	5	+1	+1	+1

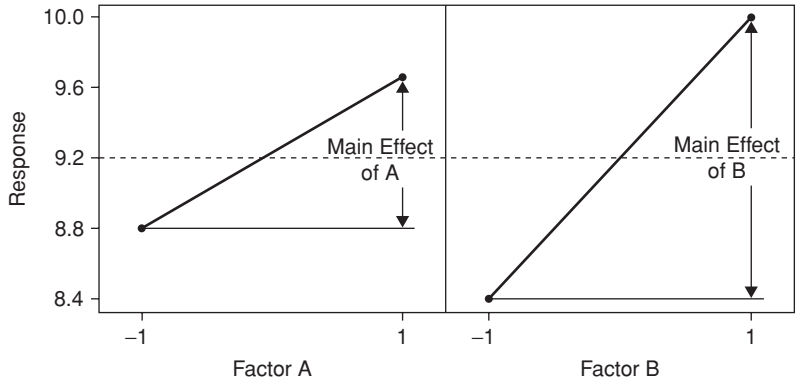


Figure 16.2 Example of effect of mutually independent factors on the response by factor A and B, respectively.

standard order of reporting facilitates explanation of the analysis. The actual execution of experimental runs should be randomized to 'distribute' the effects of any latent noise factor.

16.1.4 Main effects vs. interactions

The *main effect* of a factor is the average influence of a change in the level of that factor on the response (Figure 16.2). The *interaction effect* between two (or more) factors is the extent to which the influence of one factor on a response depends on the level of another factor (Figure 16.3). It may be synergistic or conflicting. If the interaction between two factors is significant, a focus on the interaction to identify an 'opportunity' or 'threat' takes precedence over the main effects.

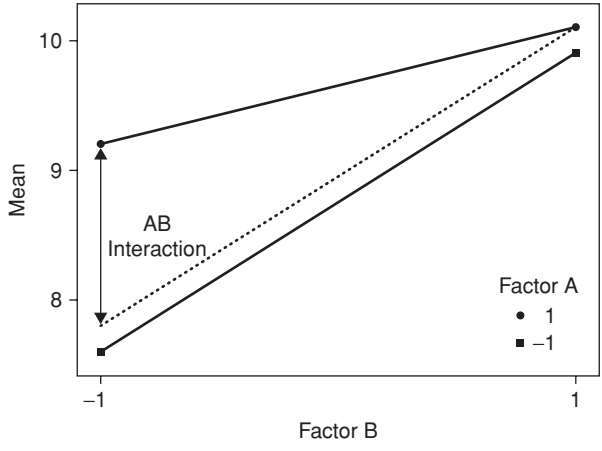


Figure 16.3 Example of effect on the response of interacting factors.

16.2 ANALYSIS OF FACTORIAL DESIGNS

16.2.1 Response function

For a 2^3 factorial design, the full model response function is defined by

$$Y = \mu + A_{i..} + B_{.j.} + C_{..k} + AB_{ij.} + AC_{i.k} + BC_{.jk} + ABC_{ijk} + \varepsilon.$$

The statistical significance of a main effect or interaction may be verified by analysis of variance (ANOVA). The ANOVA table is given in Table 16.2.

16.2.2 Reduction of model

The model may be simplified by removing main effects and/or interactions in this sequence:

(a) three-way (or higher) interactions as they are not common and/or difficult to manage in practice;
(b) two-way interactions and main effects with F statistic below unity.

$F < 1$ implies that *between* variation is less than *within* variation. If a two-way interaction is maintained, neither of the main effects may be removed. The model may then be progressively reduced by:

(a) removing effects with p-value above a defined limit (e.g. 10 %); or
(b) removing effects until the adjusted R^2 criterion shows a significant decrease.

16.2.3 Comparison and adequacy of models

For each model, determine

$$R^2 = 1 - \frac{SS_{\text{Error}}}{SS_{\text{Total}}}$$

Table 16.2 Overview of ANOVA.

Source	Degrees of freedom	Sum of square (SS)	Mean square (MS)	F statistic	p-Value
A	1	SS_A	SS_A / 1	MS_A / MS_{Error}	$F\ (F_A\,;1\,,\nu_{\text{Error}})$
B	1	SS_B	SS_B / 1	MS_B / MS_{Error}	$F\ (F_B\,;1\,,\nu_{\text{Error}})$
C	1	SS_C	SS_C / 1	MS_C / MS_{Error}	$F\ (F_C\,;1\,,\nu_{\text{Error}})$
AB	1	SS_{AB}	SS_{AB} / 1	MS_{AB} / MS_{Error}	$F\ (F_{AB}\,;1\,,\nu_{\text{Error}})$
AC	1	SS_{AC}	SS_{AC} / 1	MS_{AC} / MS_{Error}	$F\ (F_{AC}\,;1\,,\nu_{\text{Error}})$
BC	1	SS_{BC}	SS_{BC} / 1	MS_{BC} / MS_{Error}	$F\ (F_{BC}\,;1\,,\nu_{\text{Error}})$
ABC	1	SS_{ABC}	SS_{ABC} / 1	MS_{ABC} / MS_{Error}	$F\ (F_{ABC}\,;1\,,\nu_{\text{Error}})$
Error	ν_{Error}	SS_{Error}	SS_{Error} / ν_{Error}		
Total	$N-1$	SS_{Total}			

and its adjusted counterpart

$$R^2_{\text{adj}} = 1 - \frac{SS_{\text{Error}}/\nu_{\text{Error}}}{SS_{\text{Total}}/(N-1)}.$$

For comparison of models, use R^2_{adj}: the higher the value, the better the model. To check the adequacy of the 'best' model, use R^2. This metric estimates the proportion of observed variation accounted for by the model selected. For practical purposes, choose a parsimonious model with sufficient R^2.

16.3 RESIDUAL ANALYSIS

From the ANOVA table in Table 16.2, it may be observed that the significance of a main effect or interaction is dependent on MS_{Error}. Hence, it is important that we examine the distribution of the residuals. Under ANOVA, the residuals are assumed to be normally and independently distributed about a null mean and constant variance: $\varepsilon \sim \text{NID}(\mu = 0, \sigma^2 = \text{constant})$. We will examine some of the consequences when these assumptions are violated.

16.3.1 Independence

Positive autocorrelation results in underestimation of the MS_{Error}, giving rise to over-recognition of factors (Figure 16.4). The reverse is true for *negative autocorrelation*.

16.3.2 Homoskedasticity

The estimated MS_{Error} is biased towards the group with the larger subgroup size, giving rise to increased α or β (Figure 16.5).

16.3.3 Mean of zero

As shown in Figure 16.6, a trend in the residuals implies the presence of a significant predictor that has not been considered in the model.

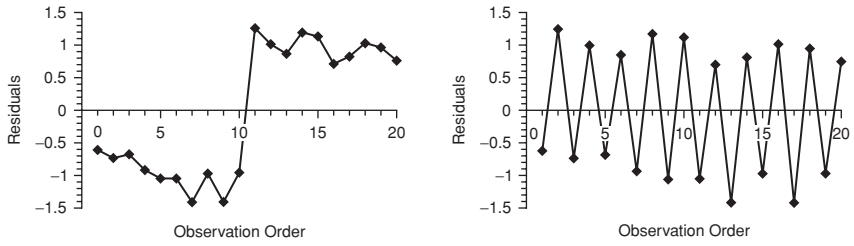


Figure 16.4 Positive (left) and negative (right) auto-correlation.

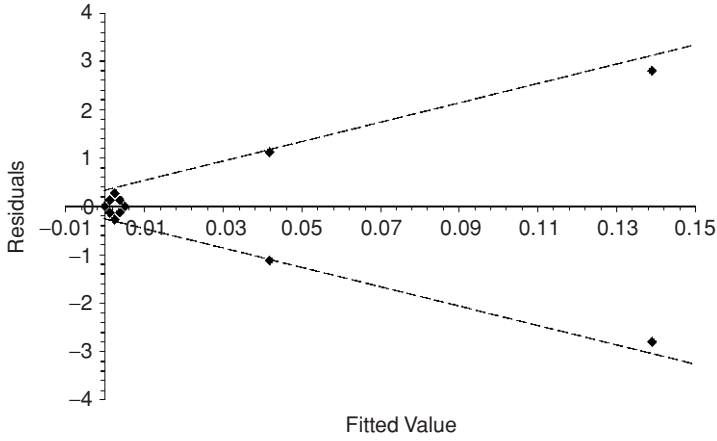


Figure 16.5 Illustration of homoscedasticity.

16.4 TYPES OF FACTORIAL EXPERIMENTS

16.4.1 Crossed designs

This is the default factorial design where each factor level is independent of the remaining factors.

16.4.2 Nested designs

Consider a 2^3 factorial experiment. Factor A is the machinery, with line 1 and line 3 as levels. Factor B is the supplies, with levels supplier A and supplier B. Factor C is the material, with levels regular and premium. As the regular grades for the two suppliers are not the same, and likewise for the premium grade, factor C is said to be

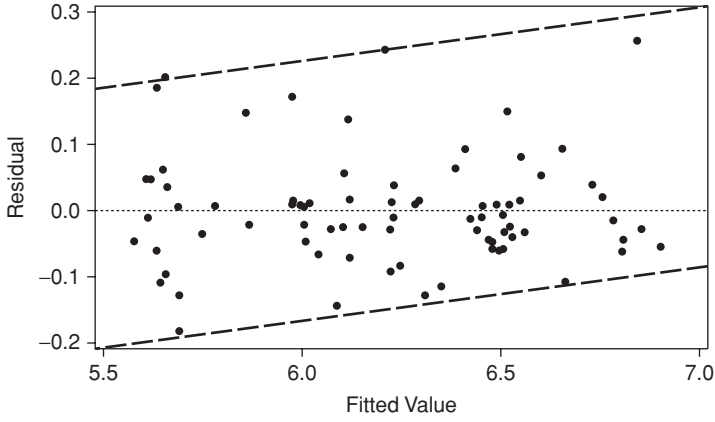


Figure 16.6 Example of a trend in residuals.

Table 16.3 Example of split-plot design.

Std order	Run order	Factor *A*	Factor *B*	Factor *C*
1	4	−1	−1	−1
2	2	+1	−1	−1
3	3	−1	+1	−1
4	1	+1	+1	−1
5	5	−1	−1	+1
6	8	+1	−1	+1
7	7	−1	+1	+1
8	6	+1	+1	+1

nested in B. Since both suppliers may be run on each of the three lines, factors A and B are said to be *crossed*. If a *nested design* is analyzed as a *crossed design*, the calculated MS_{Error} is typically higher than the actual MS_{Error}, resulting in under-recognition of the effects.

16.4.3 Split-plot designs

Consider the 2^3 factorial design in Table 16.3. Here, factor *C* is maintained at a given level (−1), while factors *A* and *B* are randomized within that level of C (runs 1–4). Factor C is then changed to the other level (+1), while factors *A* and *B* are randomized (runs 5–8).

If a *split-plot design* is analyzed as a *crossed design*, the calculated MS_{Error} is typically lower than the actual MS_{Error}, resulting in over-recognition of the effects.

16.4.4 Mixture designs

In *mixture experiments*, the product under investigation is made up of several components or ingredients. The response is a function of the proportions of the different components or ingredients. In a *crossed* or *nested design*, the response is a function of the amount of the individual components or ingredients. This is illustrated in in Figure 16.7.

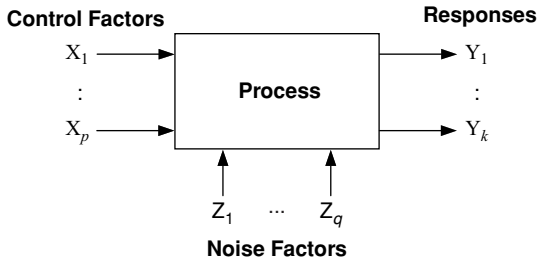


Figure 16.7 Mixture design.

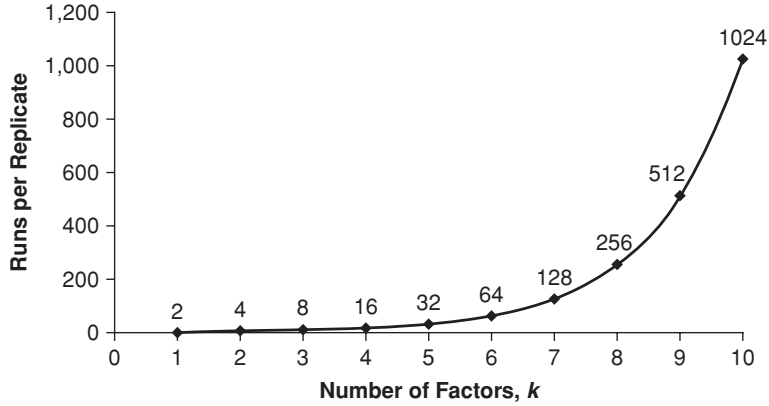


Figure 16.8 Exponential effect of number of factors on the number of runs required.

16.5 FRACTIONAL FACTORIAL DESIGNS

A *screening design* is an experimental design whose purpose is to distinguish influential factors from non-influential factors as efficiently as possible. As shown in Figure 16.8, as the number of factors increases, the number of runs per replicate of a 2^k factorial design increases exponentially.

The number of main effects and interactions that may be estimated for up to 10 factors is summarized in Table 16.4. If we limit our interests to the main effects and two-way interactions, then the required runs could be significantly reduced, as shown in Table 16.5.

A *fractional factorial design* is a subset of the factorial design. The general notation for a two-level fractional factorial design is 2^{k-p}, where k is the number of factors, and p is the *degree of fractionation* (or *number of generators*).

Table 16.4 Overview of fractional factorial design.

		Interactions								
Factors	Main effects	2-way	3-way	4-way	5-way	6-way	7-way	8-way	9-way	10-way
2	2	1								
3	3	3	1							
4	4	6	4	1						
5	5	10	10	5	1					
6	6	15	20	15	6	1				
7	7	21	35	35	21	7	1			
8	8	28	56	70	56	28	8	1		
9	9	36	84	126	126	84	36	9	1	
10	10	45	120	210	252	210	120	45	10	1

Table 16.5 Overview of fractional factorial design when only main effects and two-way interactions are of interest.

Factors	2^k design	Required runs	2^{k-p} design	Delta
2	$2^2 = 4$	3	$2^2 = 4$	
3	$2^3 = 8$	6	$2^3 = 8$	
4	$2^4 = 16$	10	$2^4 = 16$	
5	$2^5 = 32$	15	$2^4 = 16$	16 (50.00 %)
6	$2^6 = 64$	21	$2^5 = 32$	32 (50.00 %)
7	$2^7 = 128$	28	$2^5 = 32$	96 (75.00 %)
8	$2^8 = 256$	36	$2^6 = 64$	192 (75.00 %)
9	$2^9 = 512$	45	$2^6 = 64$	448 (87.50 %)
10	$2^{10} = 1024$	55	$2^6 = 64$	960 (93.75 %)

16.5.1 Confounding and aliasing

Consider the case of a 2^{3-1} fractional factorial design in Table 16.6. Closer inspection of the table should reveal that:

(a) the run order for A is identical to that for BC;
(b) the run order for B is identical to that for AC;
(c) the run order for C is identical to that for AB.

An *alias* is a factor or interaction whose pattern of levels in an experiment is identical to that of another factor or interaction. Hence, A is the alias of BC, and vice versa, while B is the alias of AC, and C that of AB.

While fractional factorial designs offer economy in terms of the number of runs required, they suffer from the inherent evil of *confounding*, that is, the influence of a factor or interaction is intermixed with that of its alias. For the above design, the influence of A and that of BC are confounded, while B is confounded with AC, and C with AB. The effect of confounding is that it adds uncertainty to the solution.

16.5.2 Example 1

Consider a 2^{3-1} fractional factorial design with the following true transfer functions:

(a) $Y_1 = \beta_1 A + \beta_2 B + \beta_3 C + \beta_{12} AB$,
(b) $Y_2 = \beta_1 A + \beta_2 B + \beta_3 C - \beta_{12} AB$,
(c) $Y_3 = \beta_1 A + \beta_2 B + \frac{1}{2}\beta_3 C + \frac{1}{2}\beta_{12} AB$.

Table 16.6 An example of a 2^{3-1} fractional factorial design.

Run	A	B	C	AB	AC	BC	ABC
1	−1	−1	+1	+1	−1	−1	+1
2	+1	−1	−1	−1	−1	+1	+1
3	−1	+1	−1	−1	+1	−1	+1
4	+1	+1	+1	+1	+1	+1	+1

Table 16.7 Result of Example 1 with zero noise.

Run	A	B	C	Y_1	Y_2	Y_3
1	−1	−1	+1	0	−2	−1
2	+1	−1	−1	−2	0	−1
3	−1	+1	−1	−2	0	−1
4	+1	+1	+1	4	2	3

Given the perfect case where noise is zero (so as to eliminate any further error in conclusion), Table 16.7 shows the results obtained. Standard analysis yields the following response functions:

(a) $Y_1 = A + B + 2C$,
(b) $Y_2 = A + B$,
(c) $Y_3 = A + B + C$.

It may be observed that when a factor and its alias are synergistic, the effect of the factor is overestimated. On the other hand, if a factor is in 'conflict' with its alias, its effect is underestimated. Hence, when a factor is found to be significant, is it the effect of the factor that is significant, the effect of the alias that is significant, or the combined effect of the factor and alias that is significant? Similarly, when a factor is found to be insignificant, is it because the factor is truly insignificant, or the factor is in conflict with its alias? In short, in the presence of confounding, we can be sure of nothing. This uncertainty grows if the AC and/or BC interactions is/are also present.

16.5.3 Example 2

A 2^{7-2} fractional factorial (1/4 fractional) design has the following alias structure:

$I + CEFG + ABCDF + ABDEG$,

$A + BCDF + BDEG + ACEFG$,

$B + ACDF + ADEG + BCEFG$,

$C + EFG + ABDF + ABCDEG$,

$D + ABCF + ABEG + CDEFG$,

$E + CFG + ABDG + ABCDEF$,

$F + CEG + ABCD + ABDEFG$,

$G + CEF + ABDE + ABCDFG$,

$AB + CDF + DEG + ABCEFG$,

$AC + BDF + AEFG + BCDEG$,

$AD + BCF + BEG + ACDEFG$,

$AE + BDG + ACFG + BCDEF$,

$AF + BCD + ACEG + BDEFG$,

$AG + BDE + ACEF + BCDFG$,

$BC + ADF + BEFG + ACDEG$,

$BD + ACF + AEG + BCDEFG$,

$BE + ADG + BCFG + ACDEF$,

$BF + ACD + BCEG + ADEFG$,

$BG + ADE + BCEF + ACDFG$,

$CD + ABF + DEFG + ABCEG$,

$CE + FG + ABCDG + ABDEF$,

$CF + EG + ABD + ABCDEFG$.

A 2^{7-3} fractional factorial (1/8 fractional) design has the following alias structure:

$A + BCE + BFG + CDG + DEF + ABCDF + ABDEG + ACEFG$,

$B + ACE + AFG + CDF + DEG + ABCDG + ABDEF + BCEFG$,

$C + ABE + ADG + BDF + EFG + ABCFG + ACDEF + BCDEG$,

$D + ACG + AEF + BCF + BEG + ABCDE + ABDFG + CDEFG$,

$E + ABC + ADF + BDG + CFG + ABEFG + ACDEG + BCDEF$,

$F + ABG + ADE + BCD + CEG + ABCEF + ACDFG + BDEFG$,

$G + ABF + ACD + BDE + CEF + ABCEG + ADEFG + BCDFG$,

$AB + CE + FG + ACDF + ADEG + BCDG + BDEF + ABCEFG$,

$AC + BE + DG + ABDF + AEFG + BCFG + CDEF + ABCDEG$,

$AD + CG + EF + ABCF + ABEG + BCDE + BDFG + ACDEFG$,

$AE + BC + DF + ABDG + ACFG + BEFG + CDEG + ABCDEF$,

$AF + BG + DE + ABCD + ACEG + BCEF + CDFG + ABDEFG$,

$AG + BF + CD + ABDE + ACEF + BCEG + DEFG + ABCDFG$,

$BD + CF + EG + ABCG + ABEF + ACDE + ADFG + BCDEFG$,

$ABD + ACF + AEG + BCG + BEF + CDE + DFG + ABCDEFG$.

With higher degrees of fractionation, there will be more confounding, more aliases for a given factor, and greater uncertainty.

Table 16.8 Design resolution and respective orders aliased effects.

Resolution	Smallest sum of order of aliased effects
III	Main effects + 2-way interactions
IV	Main effects + 3-way interactions; 2-way interactions + 2-way interactions
V	Main effects + 4-way interactions; 2-way interactions + 3-way interactions
VI	Main effects + 5-way interactions; 2-way interactions + 3-way interactions; 3-way interactions + 2-way interactions

16.5.4 Design resolution

The *resolution* of a two-level fractional factorial design is the smallest sum of the orders of aliased effects (see Table 16.8) To avoid confounding of two-way interactions, which is quite common in reality, Resolution III designs should be avoided whenever possible, while designs of Resolution V and above are preferred.

16.6 ROBUST DESIGN

16.6.1 Loss function

Quality is the loss imparted to the society from the time a product is shipped. (Genichi Taguchi)

A *loss function* is a mathematical relationship between loss (however it may be defined) and a selected performance characteristic.

16.6.1.1 Goalpost loss function

All values of a performance characteristic that lie within the specification limits are equally acceptable and, therefore, all the corresponding products are of equal quality (see Figure 16.9).

16.6.1.2 Quadratic loss function

The loss (degree of discontent) associated with a performance measure is minimized when the measure is on target, and it increases in proportion to the squared departure of the measure from its target (see Figure 16.10).

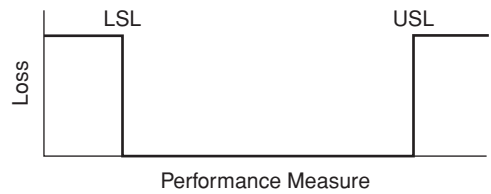


Figure 16.9 Illustration of goalpost loss function.

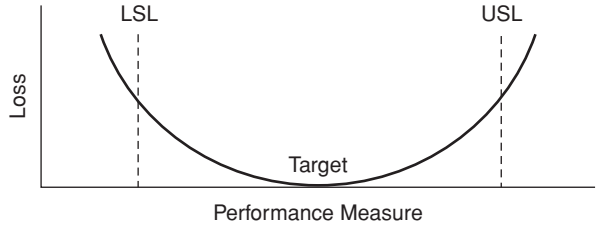


Figure 16.10 Illustration of quadratic loss function.

16.6.2 Classical DOE model vs. Taguchi DOE model

The classical design experiments model is illustrated in Figure 16.11, while the Taguchi DOE model is shown in Figure 16.12. Under the Taguchi model, appropriate treatments (combinations of *control factors*) may be determined to obtain the optimal mean (μ) and/or minimal variance (σ^2) in the response. Control factors are product design or process factors that influence a product's performance, and may be controlled by the designer or manufacturer (e.g. electrical parameters, component dimensions).

The recommended treatment should be robust enough against the effects of the *noise factors*. These are variables that affect a product's performance, but whose values cannot be controlled by the designer or manufacturer or are not controlled for economic reasons (e.g. raw material, equipment condition, labor skill, environment).

16.6.3 Taguchi's inner and outer arrays

The effects of noise factors may be included in an experimental design via *blocking*. Blocking an experiment is arranging the runs of the experiment in groups ('blocks') so that runs within each block have as much extraneous variation in common with each other as possible, – examples are using material from the same lot, evaluation under the same machine/line, and carrying out runs within a short time frame.

To accommodate both control factors and noise factors, Taguchi's design consists of two parts: an *inner array* (IA), a design involving only control factors; and an *outer array* (OA), a design involving only noise factors. Table 16.9 shows an example of a design with three control factors and three noise factors.

16.6.4 Repeats and replicates

An estimate of the *random error* (in a classical DOE model) or process variability (in a Taguchi DOE model) may be obtained by *repeated measurements* and/or *replicated runs*

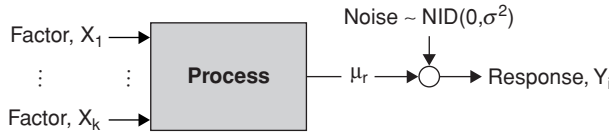


Figure 16.11 Classical DOE model.

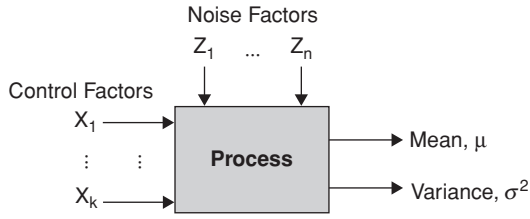


Figure 16.12 Taguchi DOE model.

(Table 16.10). Repeats are back-to-back measurements of a response at fixed factor levels. Replicates are independent runs conducted at identical sets of factor levels, in which all sources of inherent variation are present. Repeat measurements are not replicated runs because differences among such measure values do not include all sources of inherent variation.

It is not mandatory that the number of replicates equal the number of blocks required. However, the number of replicates should be sufficient to avoid confounding of the main effects and two-way interactions. Consider the example in Table 16.9 with 3 control factors and 3 noise factors (with 4 blocks). Replicates might be applied as shown in Table 16.11.

16.6.5 Determining repeats and replicates

16.6.5.1 Repeated measures

The standard deviation (and corresponding variance) determined within a run (or treatment) is the *sample standard deviation* s (and corresponding *sample variance* s^2). It

Table 16.9 A design with three control factors and three noise factors.

									OA of noise factors					
								Z_1	−1	+1	−1	+1		
								Z_2	−1	−1	+1	+1		
	IA of control factors							Z_3	+1	−1	−1	+1		
Run	A	B	C	AB	AC	BC	ABC		y_1	y_2	y_3	y_4	Mean	Variance
1	−1	−1	−1	+1	+1	+1	−1						y_1	s_1^2
2	+1	−1	−1	−1	−1	+1	+1						y_2	s_2^2
3	−1	+1	−1	−1	+1	−1	+1						y_3	s_3^2
4	+1	+1	−1	+1	−1	−1	−1		Block 1	Block 2	Block 3	Block 4	y_4	s_4^2
5	−1	−1	+1	+1	−1	−1	+1						y_5	s_5^2
6	+1	−1	+1	−1	+1	−1	−1						y_6	s_6^2
7	−1	+1	+1	−1	−1	+1	−1						y_7	s_7^2
8	+1	+1	+1	+1	+1	+1	+1						y_8	s_8^2

Table 16.10 Example of repeats and replicates runs.

Run	A	B	C	y_1	y_2	y_3	Mean	Variance
1	−1	−1	−1				y_1	s_1^2
2	+1	−1	−1				y_2	s_2^2
3	−1	+1	−1				y_3	s_3^2
4	+1	+1	−1				y_4	s_4^2
5	−1	−1	+1		1st Replicate of 3 Repeats/Run		y_5	s_5^2
6	+1	−1	+1				y_6	s_6^2
7	−1	+1	+1				y_7	s_7^2
8	+1	+1	+1				y_8	s_8^2
9	−1	−1	−1				y_9	s_9^2
10	+1	−1	−1				y_{10}	s_{10}^2
11	−1	+1	−1				y_{11}	s_{11}^2
12	+1	+1	−1				y_{12}	s_{12}^2
13	−1	−1	+1		2nd Replicate of 3 Repeats/Run		y_{13}	s_{13}^2
14	+1	−1	+1				y_{14}	s_{14}^2
15	−1	+1	+1				y_{15}	s_{15}^2
16	+1	+1	+1				y_{16}	s_{16}^2

is typically an underestimate of the *population standard deviation* s (and corresponding *population variance* s^2). For further description of the difference between the two standard deviations, refer to Chapter 6.

A practical approach towards determining the number of repeats per run is to consider the stability in the sample standard deviation. This can be achieved by examining the Shewhart c_4 constant as a function of the subgroup size (number of repeats). From Figure 16.13 it may be observed that the estimate of σ is highly unstable for subgroup size of 3 or less, but eventually stabilizes for subgroup sizes of 10 and above.

16.6.5.2 Replicated runs

The statistical significance of a main effect or interaction is verified by means of an F test (in ANOVA). For 2^k factorial (or fractional factorial) experiments, the significance is tested against an F distribution with 1 degree of freedom in the numerator and ν_e degrees of freedom in the denominator. Hence, we examine the critical values

Table 16.11 Replicates and alias structure.

Replicates	Alias structure
a) 1 replicate	Block 1 = AB; Block 2 = AC; Block 3 = BC
b) 2 replicates	Blocks = ABC
c) 3 replicates	Block 1 = AB; Block 2 = AC; Block 3 = BC
d) 4 replicates	All terms are free from aliasing

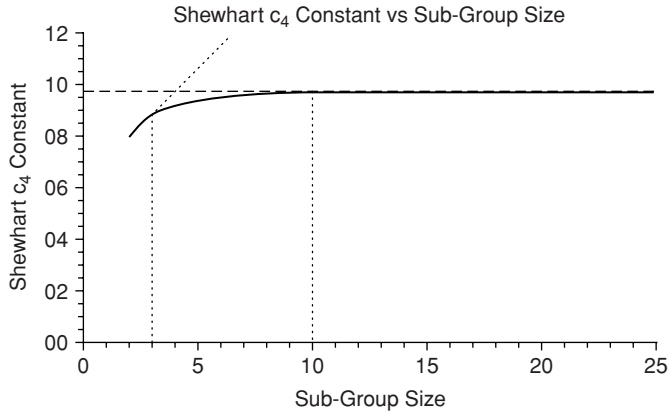


Figure 16.13 Graphical illustration of Shewhart c_4 constant vs. the subgroup size.

for $F(\alpha = 0.05, 1, \nu_e)$ distributions. From Figure 16.14 it may be observed that the significance of an effect will be difficult to prove for ν_e of 3 or less, while there is little cause for concern for ν_e of 15 and above.

In a 2^k factorial experiment, there will be k degrees of freedom for the main effects and $\frac{1}{2}k(k-1)$ degrees of freedom for the two-way interactions. With r replicates, there will be $r(2^k-1)$ total degrees of freedom for the design.

Consider the example in Table 16.9 with three control factors and three noise factors (and four blocks); the effect of replicates an confounding is shown in Table 16.12. For two replicates, the blocks are confounded with the three-way interaction ABC. If time and/or cost is a constraint and one is confident that the ABC interaction is not practically significant, then two replicates may be adequate in identifying significant

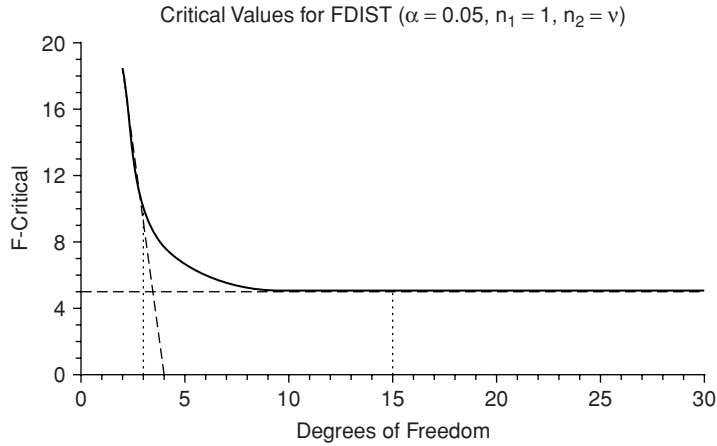


Figure 16.14 Graph of critical F-values vs. degrees of freedom.

Table 16.12 Effect of replicates on confounding.

		Number of replicates			
Source	Degrees of freedom	$r = 1$	$r = 2$	$r = 3$	$r = 4$
Main effects	k	3	3	3	3
Two-way interactions	$\frac{1}{2}k(k-1)$	3	3	3	3
Error	v_e	1	9	17	25
Total	$r\ (2^k-1)$	7	15	23	31

main effects and two-way interactions. Otherwise, four replicates would be preferred since three replicates would result in confounding between the blocks and the two-way interactions.

17

Some Strategies for Experimentation under Operational Constraints

T. N. Goh

In quality and reliability engineering, design of experiments is well established as an efficient empirical approach to characterization and optimization of products and processes. However, its actual application in industry can at times be impeded by constraints related to physical arrangements, costs, time, or irrecoverable sample damage. To enhance the feasibility and effectiveness of design of experiments, some practical strategies for overcoming such difficulties are discussed in this chapter, with an application in industrial practice as an illustration.

17.1 INTRODUCTION

Techniques of design of experiments have assumed an increasingly important role in quality and reliability engineering in recent years owing to their efficiency in data generation and the richness of results obtainable from the subsequent statistical inference.[1] In particular, two-level experiments have formed the backbone of most studies where the data requirement is greatly reduced via the use of fractional factorial designs. Applications of fractional factorial designs are further stretched when those who subscribe to Taguchi methods[2,3] routinely side-step issues of confounding of factor interactions and use mostly saturated fractional factorial designs in an attempt to include all relevant factors in a single experimental study.

This chapter is based on the article by T. N. Goh, 'Some strategies for experimentation under operational constraints', *Quality and Reliability Engineering International*, 13(5), 1997, pp. 279–283, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

With the size of experiment (number of observations) reduced to its limit, it is still commonplace in industry for an experimental design to be found impossible to implement because of operational constraints. Such concerns are addressed in this chapter, with the emphasis more on the feasibility of the suggested expedient procedures than their theoretical rigor. In the following discussions, the notation used by Box *et al.*[1] and Taguchi[2] will be used without explanation.

17.2 HANDLING INSUFFICIENT DATA

One basic requirement in the application of design of experiments is that the experimental data stipulated in the design matrix (orthogonal array) must all be available before any analysis can be initiated. Thus, in an eight-run design, regardless of whether it is a full or fractional factorial, all eight observations are needed for any assessment of the effects of the factors under study. Regular data analysis of an experiment may have to be abandoned if any of the planned observations turns out to be unavailable.

Incomplete data availability could happen for various reasons. One is technical; for example, certain combinations of factor settings may be physically difficult to arrange, or found undesirable from a safety point of view. Another is cost and time, as certain factor combinations may entail long setup time or total process time, or consumption of large amounts of test material. Yet another reason could be missing experimental data due to unexpected shortage of raw materials, equipment breakdown, mishandling of samples, or inadvertent loss of records. If the textbook data requirements cannot be fulfilled, the investigator may not be able to salvage the experimental effort. Some theoretical studies of such problems have been done in the past;[4] one related problem formulation is the inclusion, in a saturated experimental design, of more effects to be estimated than such a design would normally be able to handle.[5–8]

There is another consequence of the requirement of complete data availability. An investigator launching a large experiment, say a 2^5 factorial, while having a statistically superior design in terms of resolution of factor effects, may, in the face of industry demands for interim results, progress reports, or early decision indicators, be at a disadvantage compared to someone who uses a 2^{5-2} design in the first instance, makes some quick assessments based on the first eight data points, and goes for another follow-up 2^{5-2} experiment only when necessary. This leads to the strategy of sequential experimentation,[9] with which savings in data collection would also be possible. Much is made of the effect sparsity[10] found in industrial experiments, so that a number of effects can be reasonably assumed zero to facilitate the analysis of small experiments. This is taken one step further in lean designs[11] where, making use of part of an orthogonal design matrix, results are obtained with incomplete data sets; this is to suit situations ranging from some stop-gap, preliminary round of study to desperate damage control for an unsatisfactorily completed experiment. Against this backdrop, several variations of experimental design procedures will now be presented.

17.3 INFEASIBLE CONDITIONS

The nature of an orthogonal experimental design matrix is such that it contains a prescribed number of mandatory combinations of factor settings. Situations could arise

Table 17.1 2^{4-1} design with $4 = 123$.

I	X_1	X_2	X_3	X_4
1	−	−	−	−
2	+	−	−	+
3	−	+	−	+
4	+	+	−	−
5	−	−	+	+
6	+	−	+	−
7	−	+	+	−
8	+	+	+	+

where certain specific combinations are infeasible or best avoided, for reasons already explained. Under such circumstances, it would be useful to consider the possible variants of fractional factorials for actual use. For example, if four factors are to be studied in eight experimental runs, the recommended design is usually 2^{4-1} with $4 = 23$ as the generator, resulting in the matrix shown in Table 17.1. However, if run no. 8, say, is physically undesirable because it requires an 'all high (+)' setting for the factors, then some alternative designs can be considered, such as one shown in Table 17.2 which is based on $4 = -123$. If, on further planning, it is felt that one particular combination of factor settings, say no. 7 in Table 17.2, is operationally difficult to arrange, then yet other designs can be used to avoid the $(-, +, +, +)$ requirement; one such design is shown in Table 17.3, which is based on $4 = -23$.

In some other situations, data may already exist for a particular factor setting combination, so it would be desirable to include that combination in the design. For example, if $(+, -, +, +)$ is to be included, whereas $(+, +, -, +)$ is to be excluded, then a matrix generated by $4 = -12$ would be appropriate. In all, as many as eight alternative 2^{4-1} designs are available for the study of four factors in eight experimental runs, based on generators $4 = 123, 4 = -123, 4 = 12, 4 = -12, 4 = 13, 4 = -13, 4 = 23$ and $4 = -23$. The structures of design matrices based on these generators are summarized under columns A, B, C, D, E, F, G and H respectively in Table 17.4, showing

Table 17.2 2^{4-1} design with $4 = -123$.

I	X_1	X_2	X_3	X_4
1	−	−	−	+
2	+	−	−	−
3	−	+	−	−
4	+	+	−	+
5	−	−	+	−
6	+	−	+	+
7	−	+	+	+
8	+	+	+	−

Table 17.3 2^{4-1} design with $4 = -23$.

I	X_1	X_2	X_3	X_4
1	−	−	−	−
2	+	−	−	−
3	−	+	−	+
4	+	+	−	+
5	−	−	+	+
6	+	−	+	+
7	−	+	+	−
8	+	+	+	−

clearly which eight out of the 16 experimental runs of a full 2^4 factorial are used for a particular half-fraction design. Thus, columns A, B, H correspond to Tables 17.1, 17.2, and 17.3, respectively. Such a master table of 2^{4-1} designs can be conveniently used to expressly exclude or include certain factor setting combinations; in many cases, more than one operational constraint or requirement can be accommodated simultaneously.

17.4 VARIANTS OF TAGUCHI ORTHOGONAL ARRAYS

The flexibility afforded by the experimental design approach illustrated above is absent from the usual Taguchi routine which expects an experimenter to obtain his design from one of the standard orthogonal arrays appended at the end of a Taguchi

Table 17.4 Alternative 2^{4-1} designs out of a full design.

	X_1	X_2	X_3	X_4	A	B	C	D	E	F	G	H
1	−	−	−	−	1			1		1		1
2	+	−	−	−		2	2		2			2
3	−	+	−	−		3	3			3	3	
4	+	+	−	−	4			4	4		4	
5	−	−	+	−		5		5	5		5	
6	+	−	+	−	6		6			6	6	
7	−	+	+	−	7		7		7			7
8	+	+	+	−		8		8		8		8
9	−	−	−	+		1	1		1		1	
10	+	−	−	+	2			2		2	2	
11	−	+	−	+	3			3	3			3
12	+	+	−	+		4	4			4		4
13	−	−	+	+	5		5			5		5
14	+	−	+	+		6		6	6			6
15	−	+	+	+		7		7		7	7	
16	+	+	+	+	8		8		8		8	

Table 17.5 Taguchi textbook $L_9(3^4)$ design.

I	x_1	x_2	x_3	x_4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

textbook. Users of Taguchi methods in industry who do not possess the necessary background in statistics may not be aware of the fact that Taguchi orthogonal arrays are by no means unique; thus an alternative $L_4(2^3)$ array is possible with the use of a $3 = -12$ generator, and many more possibilities exist for the construction of an $L_8(2^7)$ array with the use of negative generators. The $L_9(3^4)$ array is similarly nonunique; there is no reason why one should be confined to the particular design, shown in Table 17.5, offered in all Taguchi methods publications. If, for example, in a product development study, the (1, 1, 1, 1) combination is deemed undesirable or difficult to implement, whereas there are existing prototypes featuring (2, 2, 2, 2), then one can redesign an orthogonal array such as that shown in Table 17.6 where these exclusion and inclusion requirements are satisfied.

To fully exploit and benefit from such flexibility in experimental design, it is highly recommended that experimental designers be familiar with the construction of Latin squares (and hence Greco-Latin and higher structures). The technique involved is in fact rather simple: *in a balanced design, every level of every factor is associated with every level of every other factor*. With this basic principle, a project in design of experiments would truly begin with a *design* exercise, entailing the development of a matrix which can accommodate requirements and meet constraints to the extent allowed by the principle.

Table 17.6 Alternative $L_9(3^4)$ design.

I	x_1	x_2	x_3	x_4
1	1	1	3	2
2	1	2	1	3
3	1	3	2	1
4	2	1	1	1
5	2	2	2	2
6	2	3	3	3
7	3	1	2	3
8	3	2	3	1
9	3	3	1	2

17.5 INCOMPLETE EXPERIMENTAL DATA

The experimental design process could also encounter cases where no satisfactory matrix can be constructed to exclude all physically infeasible conditions. Under such circumstances, unless the experiment is abandoned altogether, the resulting data set will not be complete. In certain other situations, the loss of essential measurements is unexpected, owing to spoilage of samples or sudden equipment breakdown. Regardless of whether the missing data is foreseeable, it is still possible to extract whatever information is available in the incomplete data set via an expedient known as 'lean design',[11] making use of the effect sparsity principle: since not all effects generated by the factors in an experiment are significant, one may depend on reduced degrees of freedom in the data set to perform the necessary statistical analysis after some a priori assignment of zero values to certain effects. In such an event, experience and technical judgment will be extremely useful in determining which effects to ignore in order to carry out a remedial data analysis.

17.6 ACCURACY OF LEAN DESIGN ANALYSIS

It is useful to note that the accuracy of the above analysis depends only on the validity of the zero-value assumption about the selected effects, and is unrelated to the value of any particular missing data. For an illustration of this principle, consider a 2^{7-4} design with $4 = 12$, $5 = 13$, $6 = 23$ and $7 = 123$ as generators (Table 17.7). Let an element of the design matrix in row i and column j be denoted x_{ij}, $i = 1, 2, \ldots, 8$, $j = 1, 2, \ldots, 7$; a given x_{ij} is either -1 or $+1$. The effect of factor j is given by

$$E_j = \frac{1}{4} \sum_i (x_{ij}) y_i . \tag{17.1}$$

Suppose for some reason that y_m, $1 \leq m \leq 8$, is unavailable, and based on prior knowledge or experience, E_n, $1 \leq n \leq 7$, is judged the least likely to be significant. Since

$$E_n = \frac{1}{4} \sum_i (x_{in})\, y_i \tag{17.2}$$

Table 17.7 A 2^{7-4} design.

I	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_i
1	−	−	−	+	+	+	−	y_1
2	+	−	−	−	−	+	+	y_2
3	−	+	−	−	+	−	+	y_3
4	+	+	−	+	−	−	−	y_4
5	−	−	+	+	−	−	+	y_5
6	+	−	+	−	+	−	−	y_6
7	−	+	+	−	−	+	−	y_7
8	+	+	+	+	+	+	+	y_8

the assumption translates into

$$\frac{1}{4}\sum_{i}(x_{in})\,y_i = 0 \tag{17.3}$$

From equation (18.3), y_m can be solved to give its 'dummy' value

$$\hat{y}_m = -\sum_{i\neq m}(x_{in})\,y_j. \tag{17.4}$$

Thus, with reference to (18.1), all the effects in this experiment can now be estimated:

$$\hat{E}_j = \frac{1}{4}\sum_{i\neq m}(x_{ij})\,y_i + \frac{1}{4}(x_{mj})\,\hat{y}_m \tag{17.5}$$

where E_j, $\hat{E}_j$are the values of the effect of the jth factor calculated based on a set of eight actual measurements and of seven actual measurements, respectively. With (18.1), (18.2), and (18.5), it can be shown[12] that the error associated with the estimate based on the reduced measurement set is

$$E_j - \hat{E}_j = \frac{x_{mj}}{x_{m_n}}E_n. \tag{17.6}$$

Since the error is independent of the values of both y_m and $\hat{y}_m$, the validity of the zero-effect assumption about the nth factor is far more important than the fact that y_m has not been presented for the data analysis. This is a reassuring mathematical conclusion in view of the fact that it is often easier to judge which factor is insignificant than to hazard a guess at a value for a missing experimental measurement.

17.7 A NUMERICAL ILLUSTRATION

The above analysis will now be illustrated by a numerical example. In a product reliability study, the failure of an electronic device was traced to weaknesses of an ultrasonic weld in the device housing. Five factors of the welding process were included in a troubleshooting experiment. Owing to the high cost of the devices, a 2^{5-2} screening design was used, with the coded design and results exhibited in Table 17.8. The design matrix is essentially the same as that shown in Table 17.7, with x_1, x_2, x_3, x_4 and

Table 17.8 2^{5-2} screening experiment for a welding process.

Air pressure	Hold time	Down speed	Weld time	Trigger pressure	x_5	x_6	Weld strength
−	−	−	+	−	+	+	27.9
+	−	−	−	+	−	+	17.3
−	+	−	−	+	+	−	17.1
+	+	−	+	−	−	−	39.1
−	−	+	+	+	−	−	36.5
+	−	+	−	−	+	−	18.6
−	+	+	−	−	−	+	16.0
+	+	+	+	+	+	+	34.3

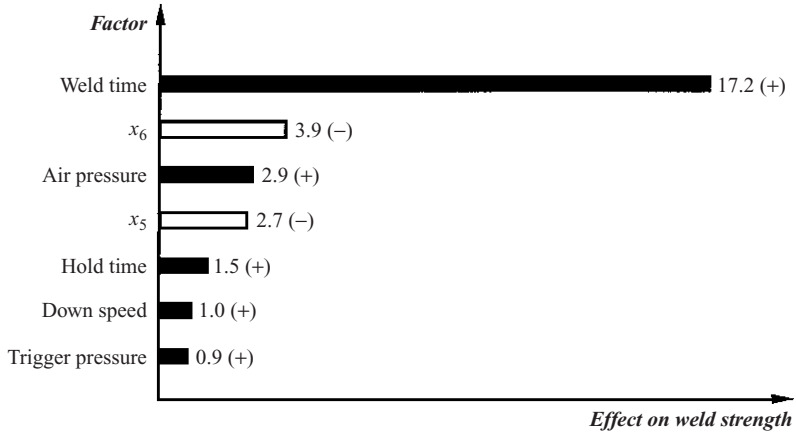


Figure 17.1 Results of a 2^{5-2} welding experiment.

x_7 assigned to air pressure, hold time, down speed, weld time and trigger pressure, respectively. With the full data set, the effects of the five factors are summarized in a Pareto plot in Figure 17.1. It is evident that even allowing for the existence of some interaction effects, weld time stands out as the most critical factor in affecting weld strength. The problem was subsequently solved via an upward adjustment of weld time from the prevalent set point.

With these results, it could be reasoned that had the investigators been requested to reduce the amount of destructive tests with the high-cost samples, one of the eight experimental runs could perhaps have been omitted. For example, if the last run were not performed, then either one of the effects associated with x_5 and x_6 could be assumed zero, and, by virtue of (18.6), each of the estimated effects of the five factors would be altered by 2.7 and 3.9, respectively. This would not affect the conclusions of this screening experiment at all, since with the new numerical results, weld time would still reveal itself as the predominant factor in a Pareto analysis. In such a situation, the quality of decisions made in a screening experiment is seen to depend much more on the relative strengths of the factor effects than on the completeness of the experimental data set.

17.8 CONCLUDING REMARKS

Inasmuch as textbook conditions cannot be always satisfied in industrial situations, planning of experiments should be done more in the spirit of adaptive design rather than by unimaginative adoption of recipes. It should be pointed out that an ability to deviate from the standard orthogonal arrays of Taguchi textbooks would considerably broaden the application of Taguchi methods, a fact yet to be recognized by many of the methods' advocates. Non-Taguchi types who know enough to evolve their own fractional factorial designs would appreciate the possibility of 'customizing' a design matrix to cope with real-life constraints. Coupled with technical judgment, suitable sequential and incremental experimental design and analysis schemes[9–11] can help

an investigator overcome operational obstacles in applying design of experiments. Admittedly, there are inherent risks associated with the judgments and assumptions made along the way; but as in investments in the financial world, the returns in the end tend to be proportional to the risks a person is willing or able to take in making his decisions. What is important, in both financial investment and industrial design of experiments, is a keen awareness of the possible risks involved, and an ability to balance such risks against potential returns. Again, in both cases, no textbook rules can replace insight and experience in the course of optimizing the final outcome.

REFERENCES

1. Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978) *Statistics for Experimenters*. New York: John Wiley & Sons Inc.
2. Taguchi G. (1986) *Introduction to Quality Engineering*. Tokyo: Asian Productivity Organization.
3. Goh T.N. (1993) Taguchi methods: some technical, cultural and pedagogical perspectives. *Quality and Reliability Engineering International*, 9, 185–202. See also Chapter 18, this volume.
4. Cheng, C.S. and Li, C.C. (1993) Constructing orthogonal fractional factorial designs when some factor-level combinations are debarred. *Technometrics*, 35, 277–283.
5. Booth, K.H.V. and Cox, D.R. (1962) Some systematic supersaturated designs. *Technometrics*, 4, 489–495.
6. Franklin, M.F. (1985) Selecting defining contrasts and confounded effects in *pn-m* factorial experiments. *Technometrics,* 27, 165–172.
7. Lin, D.K.J. (1993) A new class of supersaturated designs. *Technometrics*, 35, 28–31.
8. Wu, C.F.J. (1993) Construction of supersaturated designs through partially aliased interactions. *Biometrika*, 80, 661–669.
9. Box, G.E.P. (1993) Sequential experimentation and sequential assembly of designs. *Quality Engineering*, 5, 321–330.
10. Box, G.E.P. and Meyer, R.D. (1986) An analysis of unreplicated fractional factorials. *Technometrics*, 28, 11–18.
11. Goh T.N. (1996) Economical experimentation via 'lean design'. *Quality and Reliability Engineering International*, 12, 383–388. See also Chapter 19, this volume.
12. Goh T.N. (1997) Use of dummy values in analyzing incomplete experimental design data. *Quality Engineering*, 10, 397–401.

18

Taguchi Methods: Some Technical, Cultural and Pedagogical Perspectives

T. N. Goh

It has been a decade since the term 'Taguchi methods' became part of the vocabulary of the quality profession in the West, particularly the USA. The decade has also seen heated debates conducted by two distinct camps of professionals, one unfailingly extolling the new-found virtues and power of Taguchi methods, and the other persistently exposing the flaws and limitations inherent in them. Against this backdrop, this chapter offers some pertinent perspectives on the subject for those who have to decide 'to Taguchi, or not to Taguchi' in personnel training and actual applications. In the final analysis, as for most issues, the middle ground seems to be the most sensible way out; however, the chosen course of action will be on much firmer ground if it is based on a good understanding of the interplay of factors underlying the controversies – technical, cultural, and even pedagogical. For this reason such factors are highlighted in this chapter in a non-mathematical language for the benefit of decision makers and quality practitioners in industry who cannot afford the time to wade through the theoretical discourses in the literature, although academics may find some of the views expressed here subjects for further discussion and research in themselves.

This chapter is based on the article by T. N. Goh, 'Taguchi methods: Some technical, cultural and pedagogical perspectives', *Quality and Reliability Engineering International*, 9, 1993, pp. 185–202, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

18.1 INTRODUCTION

A little over a decade ago, Genichi Taguchi, a Japanese engineer, spent some time at AT&T Bell Laboratories in New Jersey, and demonstrated how statistical design of experiments could enhance the efficiency of empirical investigations in industrial research and development, rationalize the product realization process, and improve the quality and reliability of manufactured products. Before long, a collection of ideas and techniques under the label 'Taguchi methods' began to spread to other organizations, and began to be hailed by their advocates as the long-awaited cure for the faltering quality performance and declining competitiveness of American manufacturing industries. However, the Taguchi movement was soon met with criticisms and challenges from many veterans in the quality field who, led by academics and applied statisticians, subjected the various Taguchi procedures to close scrutiny; not a few of them have warned against 'quick and dirty' solutions to quality problems and advised those seeking valid answers to depend more on methodologies built upon firmer theoretical grounds.

The debate rages on today, although the different schools of thought have already been well aired in seminars, conferences, journals and books.[1–20] Although descriptions and commentaries had initially appeared in a variety of non-academic publications,[21–24] the paper by Kacker and the accompanying discussion[8] marked the first formal attempt to introduce Taguchi methods in a major journal. After several years of debate in the USA, Box gave some comprehensive expositions of applied statisticians' views.[14,15] More recently, Pignatiello and Ramberg[19] summarized the pros and cons of Taguchi methods for those who are familiar with the details, and Nair and his panel of discussants[20] went deeper into the academic aspects of 'parameter design', the main pillar of Taguchi methods. Studies of Taguchi methods and related matters can be expected to continue for years to come.

Management and technical decision makers in industry, who may only have a cursory exposure to the subject but little inclination for procedural details or academic debates, will have a need at this juncture to gain some essential insights into the issues involved, so as to be able to make more informed judgments on the viewpoints and techniques that they face from time to time. To this end this chapter presents, in a non-mathematical language, an overview and analysis of Taguchi methods from several new angles, encompassing technical, cultural, and pedagogical considerations: it is not an attempt to arrive at definitive conclusions, but is meant to add further perspectives to this important and practical subject.

18.2 GENERAL APPROACHES TO QUALITY

There are several dimensions to the quality improvement initiatives of an organization. The most fundamental is a properly established quality management system, which encompasses subjects ranging from corporate philosophy to policies, procedures, employee motivation, and even supplier management and customer relations, and can be typified by the ISO 9000 series descriptions. This should be complemented by the requisite quality technology, entailing engineering capabilities and resources commensurate with the technical performance required of the products or services

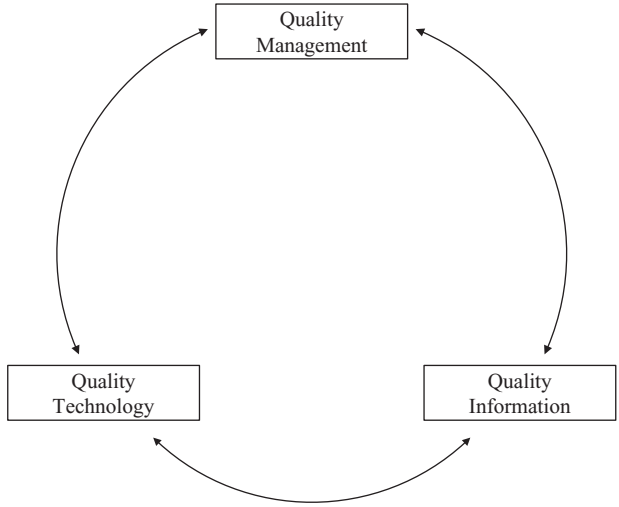


Figure 18.1 Important dimensions of quality improvement initiatives.

to be generated. The full quality and productivity potential of an organization can be realized, however, only with the presence of a third dimension, namely a quality information capability. Quality information know-how is statistical in nature; it is essential for intelligent decision making based on data collected from the performance of processes or products in the face of natural variabilities. Statistical quality control (SQC)[25] is an established area of study dedicated to this requirement; as will be discussed in the next section, there is now a wider scope of application of statistical techniques for quality improvement beyond traditional SQC. Generally, only an effective interplay of management, technology, and information can bring about balanced and sustained advances in quality levels, as depicted in Figure 18.1.

18.3 STAGES IN STATISTICAL APPLICATIONS

In chronological terms, the adoption of statistical techniques in manufacturing industries has progressed in three broad stages. The first is product inspection, where statistical sampling plans help determine the product sample size and decision rules (e.g. the acceptance number). Sampling inspection is, strictly speaking, not a quality improvement tool; all it does is attempt to detect products not conforming to requirements during inspection, the effectiveness of the attempt being reflected by operating characteristic (OC) curves of sampling plans. Its futility is expressed in the oft-repeated saying: 'Quality cannot be inspected into the product'.

At the next stage, attention is turned 'upstream', that is, to the process that generates the product in question, leading to techniques such as process capability studies and process control chart applications. The effectiveness of statistical process control (SPC) lies in its ability to prevent the generation of unsatisfactory products; however, as in acceptance sampling, this is basically a negative and passive approach, since no

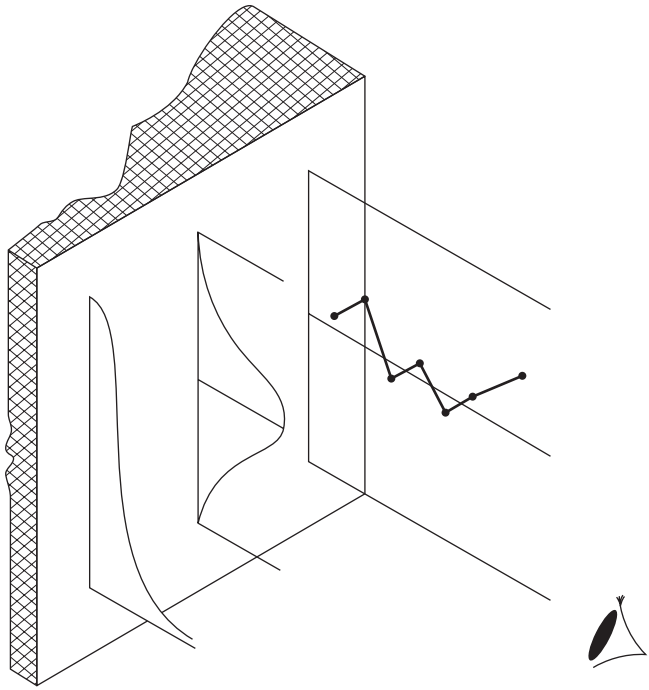


Figure 18.2 Traditional modes of statistical quality control: focus at the output end.

attempt is made to change the process for the better. Even when a process is under control, say as indicated by a p chart (for product proportion defective), there is no evidence at all that the proportion defective reflected by the centerline of the chart is the best that one can get out of the process – in other words, that the process has been fine-tuned to its optimal state. Figure 18.2 illustrates the fact that product inspection and process control are mere monitoring activities: OC curves, process capability studies, control charts, and the like all deal with the *output end* of a product or process.

It is clear, then, that to fundamentally remove the possibility of defective products, or to obtain the best performance of a given process, one has to get to the input end of the process, manipulate and fine-tune the input variables – temperature, pressure, time, raw material property, and so on – in such away that the output is optimized, where output refers to one or more measurable performance indexes such as yield, defective rate, or some quality characteristic such as length, voltage, or power. Such reasoning is straightforward, but adjustment of input variables in practice has long been based on equipment makers' recommendations, or experience, judgment, even trial and error on the part of technical personnel. This is because there is usually a lack of knowledge concerning the linkage between the input (x_1, x_2, ...) and output (say two characteristics y_{I} and y_{II}) in a given product or process (the black-box P) as depicted in Figure 18.3 or, expressed mathematically,

$$y_{\mathrm{I}} = f(x_1, x_2, \ldots, x_k), \qquad y_{\mathrm{II}} = g(x_1, x_2, \ldots, x_k), \tag{18.1}$$

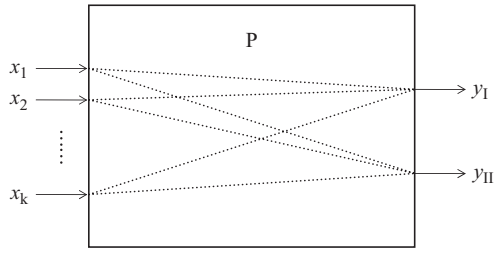


Figure 18.3 Input–output linkage of a product or process.

owing to the complexity of realistic processes and products in which a considerable number of input variables are usually involved (k such variables in Figure 18.3).

While derivation of equations (18.1) from first principles of physical sciences is all but impossible, attempts to obtain them by empirical means could prove unproductive, as most technical personnel would conduct the study with the traditional 'one variable at a time' procedure of experimentation. Such a procedure not only entails a large number of observations, but would also fail to bring out interactions among the variables x_1, x_2, ... and does not take into account the effects of noise variables (undesirable disturbances) that have not been explicitly singled out for examination.

The problem of empirically obtaining valid input–output relations, or mathematical models, was actually handled more than half a century ago by agricultural researchers led by Fisher.[26,27] The problems they faced were similar in nature: they had to discover the linkage between the yields of crops and settings of manipulable inputs such as amount of irrigation, type of fertilizer, soil composition and so on, for which no quantitative cause-and-effect relationships could be derived theoretically. Fisher's methodology, known as *design of experiments*, discards the 'one variable at a time' concept and enables the investigator to make use of only a small number of experimental data to disentangle the effect of each input variable on the output, isolate the interactions that may exist, and explicitly assess the noise effects in the physical phenomenon under study. Understanding of complex input–output relations via empirical investigations thus became feasible.

The potential of design of experiments remained largely untapped by industry till after the Second World War.[27,28] Subsequently, applied statisticians, represented most notably by George E.P. Box, William G. Hunter and J. Stuart Hunter who also authored a seminal work,[29] started to educate engineers in statistical design of experiments, in addition to SPC, for quality improvement purposes. They helped bring out the third stage of advancement in the industrial application of statistics: the objective now is to pre-empt the occurrence of defective products, not just detect or prevent it; an active rather than passive approach is advocated in process management, and strategies such as response surface methodology[30,31] and evolutionary operation[32] cap the effort to optimize the performance of black-box systems. Partly owing to the statistical language used in the presentation of techniques, industries in the West, except certain large chemical engineering companies, did not readily adopt design of experiments on a large scale before the 1980s. Table 18.1 summarizes the three broad stages that characterize the advances in statistical applications in industry.

Table 18.1 Advances in the application of statistics for quality improvement.

Stage	I	II	III
Philosophy	Quality by inspection	Quality by monitoring	Quality by design
Strategy	Passive	Defensive	Pre-emptive
Approach	Defect detection	Defect prevention	Defect elimination
Objective	Damage control	Status quo	Optimization
Technique	Sampling inspection	Process control	Robust design
Tools	Sampling plans	Indexes and charts	Design of experiments
Basis	Data based	Information based	Knowledge based
Location	Check point	On-line	Off-line
Application	Product	Process	Product and process
Mode	Batch by batch	Line by line	Project by project
Currency	1940s	1960s	1980s

18.4 THE TAGUCHI APPROACH

Before Taguchi methods spread throughout the USA, design of experiments and the associated techniques were treated as mathematical tools, more like an adjunct to an engineer's technical resources for the study of product and process characteristics. In fact, they were taught and studied largely in this manner both in the universities and in continuing education courses for engineers. Parallel to this, in the pre-Taguchi period, was a parade of personalities in the USA preaching a variety of philosophies, viewpoints, and strategies for quality – Deming, Juran, Crosby and Feigenbaum being some of the better-known names. The target audience of such quality 'gurus' was mostly senior management; engineers who were responsible for design and manufacturing were largely left to their own devices – mostly technological – to raise quality performance. Thus with reference to Figure 18.1, many organizations during this period were operating with a disjointed rather than integrated front for quality improvement which, in the face of competition from superior Japanese quality performance, constituted a backdrop to the ultimate cry in the summer of 1980: 'If Japan can, why can't we?' (this was also the title of a much publicized NBC television documentary).

A partial answer to the question can actually be glimpsed from Taguchi methods. Although 'Taguchi methods' is a term coined by Americans for the convenience of Americans, it does serve as a convenient reference to the approach with which the Japanese have been able to harmonize and integrate management, technology, and information capabilities to enhance quality, reliability, and profitability in manufactured products. Some relevant background is as follows.

After the War, Genichi Taguchi used and promoted in Japan statistical techniques for quality from an engineer's perspective, rather than that of a statistician. The starting point in Taguchi methods is his unconventional definition of quality. In contrast to concepts such as 'fitness for use', 'conformance to requirements', and 'customer satisfaction' circulated in the West, Taguchi's 'loss to society' definition reflects two common oriental values, namely aspiration to perfectionism and working for the collective good. Thus, quality is not cast as a one-off performance orchestrated to suit an

individual, isolated 'customer' (which could range from an actual human being to 'the next process' in a manufacturing system) at any one time, but is regarded as something that must be made so pervasive – much like working telephones and electric lighting in a modern office – that its existence is felt only through its absence. Once quality is not achieved, every party in an interlocking relationship suffers: suppliers, manufacturers, customers and so on, an entire chain of elements in society. Indeed, 'loss to society' reflects such a lofty ideal that it transcends all lesser visions of quality, and remains the only major thought of Taguchi's that has remained largely unscathed in assaults on Taguchi methods through the years.

'Loss to society' is next measured through the deviation of actual product quality from its target value. Use is made of a 'loss function', a mathematical expression that can state, particularly for management purposes, the monetary value of the consequence of any shortfall (or improvement) in quality. Although such a monetary value more often than not represents only a virtual loss or gain, it is a convenient performance index that can be readily appreciated by decision makers – a very important feature played up by Taguchi methods promoters to draw the interest of managers. This is in sharp contrast to the presentations of statisticians which traditionally cannot be divorced from the language of mathematical formalism and abstraction: hypotheses, alpha and beta errors, white noise, OC curves and such like.

To management and technical personnel, the Taguchi's approach to quality has another obvious appeal that is not seen in the established SQC literature. This lies in the way Taguchi admonishes them to leapfrog the earlier stages of statistical applications shown in Table 18.1 and strive to improve quality at the out set during product and process design and development, making use of design of experiments for what he calls 'off-line quality control'.[3] Thus 'Quality must be designed into the product' is not a mere slogan; Taguchi wants to have it actually carried out to provide a fundamental solution to possible quality problems. Basically this is to be achieved through the understanding and exploitation of product and process behavior as expressed in equation (18.1), which is a motivation shared by both Taguchi and mainstream statisticians, but is also where Taguchi methods begin to depart from statisticians' formulation of the physical system in question. The next section will highlight some technical ideas that have helped shape the uniqueness of Taguchi methods, yet have given rise to many of the controversies one has seen.

18.5 TAGUCHI'S 'STATISTICAL ENGINEERING'

As Taguchi's techniques for product and process optimization have already been well documented and interpreted,[9] an account of procedural matters will not be necessary here. It may be fairly stated that these techniques are invariably motivated by engineering considerations, using statistical tools to serve engineering purposes – in contrast to statisticians' techniques that are primarily mathematically formulated, then subsequently exploited for their potential to solve engineering problems. Thus Taguchi's approach may be regarded as 'statistical engineering' rather than 'engineering statistics', the latter being a term traditionally used by statisticians to cover mathematical techniques extracted for learning and hopefully application by engineers. The main statistical contents of Taguchi methods may be appreciated from the following: problem

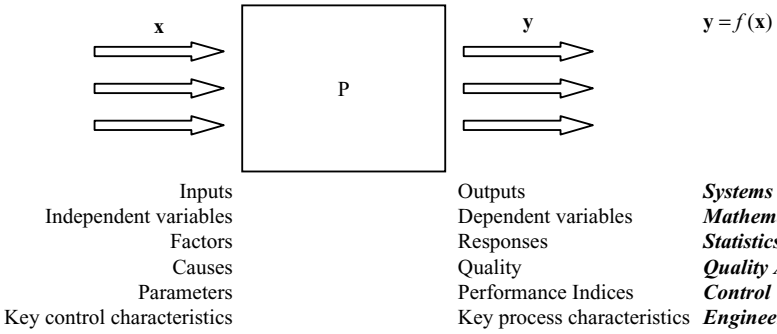


Figure 18.4 Framework of product or process performance optimization.

formulation, experimental design, data analysis, special applications, and finally some illustrative examples.

18.5.1 Problem formulation

The common starting point for all studies is the framework shown in Figure 18.4 in which the black box P represents either a product or process, and the various ways in which input and output variables have been labeled by people in different fields are shown. It is desired to determine the set-points for the manipulable input parameters to suit quality objectives exhibited by the resultant output performance. For consistency, from now on the terms *parameter* and *response* will be used, and only a single response will be considered in the discussion as extension to multiple responses is straightforward.

The statisticians' model of the problem under study is illustrated in Figure 18.5. The product or process behavior is represented by

$$\begin{aligned} y &= \hat{y} + e \\ &= f(x_1, x_2, x_3) + e, \end{aligned} \tag{18.2}$$

where y is the response that reflects performance, x_1, x_2, x_3 are controlled parameter values, and e a random variation in y, known as *error* or simply *noise*, which reflects variations in the response attributable to uncontrollable or unknown sources such as environmental conditions. Statistically this variation is assumed to be normally

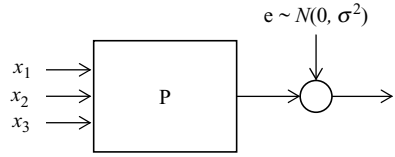


Figure 18.5 Traditional model of noise effects on response.

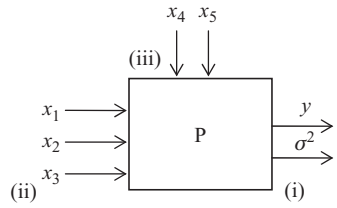


Figure 18.6 Taguchi's model of noise effects.

distributed, has no long-term systematic value, but is characterized by a constant variance:

$$e \sim N\left(0, \sigma^2\right). \tag{18.3}$$

Hence y is subject to a constant variance σ^2 whenever the values of parameters x_1, x_2, x_3 are fixed.

Such a model has been used by statisticians for design of experiments for a long time, and formed the basis of work by Box and others[28–35] before the advent of Taguchi methods. Taguchi's model is typified in Figure 18.6. It differs from the previous one in at least three ways: (i) The variation in response, reflected by σ^2, is not constant; instead, it is considered a function of *design parameters* x_1, x_2, x_3. (ii) Such parameters *do not necessarily* stay at the set (nominal) values during the actual operation of a product or process as is desired and commonly assumed. (iii) Certain external noise parameters (say x_4, x_5) can be, where possible (e.g. by simulation), included in the study of design parameters (x_1, x_2, x_3) so that the latter can be optimized with a view to counteracting the effects of the former .

From an engineering point of view, the above features do add realism to problem formulation, as (i) consistency, not just the average, of response can sometimes be directly influenced by design choice, such as the amount of pressure used in an assembly process, or the type of material used in a product; (ii) the nominal value recommended for a product or process is not necessarily what is realized; for example, there is always piece-to-piece variation of the actual resistance value of a resistor specified for a product and, in the case of a process, what is actually found in an oven often does not correspond to the nominal temperature setting; (iii) parameters such as environmental temperature and humidity can and should be simulated at the design and development stage to expose possible weaknesses in a product or process; ways should then be found to improve both the level and variation of y through judicious adjustments of design parameters.

It is useful to note that the impact of noise is traditionally regarded as inevitable but uniform in statistical design of experiments, which is a reasonable assumption in agricultural experiments because weather and other natural elements do exert a consistent influence over all the plants under study. Taguchi, however, treats variation in response as a subject of study in itself: this again is understandable because noise is the very source of quality and reliability problems; not only should it be recognized and assessed, but also efforts should be made to remove or neutralize it where possible. This idea then leads to yet another difference between Taguchi and mainstream statisticians, in the way experiments are designed.

18.5.2 Experimental design

Generally, the power of experimental design lies in its ability to: (i) study cause-and-effect relations between design parameters and response with only a small amount of experimental data; (ii) isolate and quantify the effect of each parameter on the response; (iii) study the interaction effects among the various parameters; and (iv) assess the statistical significance of the quantified effects. Taguchi packaged the requisite studies into what he termed *parameter design* to achieve the above objectives, and had them reduced to step-by-step routines simple enough to be implemented by those without any background in statistics. Parameter design is then presented as an intermediate stage between *system design* (based on engineering knowledge and technical know-how) and *tolerance design* (which defines tolerances for manufacturing). Many learners, however, have regarded parameter design as the essence of Taguchi methods.

The result of Taguchi's packaging of experimental design is a set of 'cookbook' procedures that involve the selection of a pattern of parameter settings from a standard set of experimental design matrices known as *orthogonal arrays*, using a graphical aid called linear graphs.[1–3] The user would not know the rationale behind the manipulation of parameter settings executed during an experimental study, nor would he be able to alter the designs to meet any practical constraints faced in the study. For example, Taguchi's $L_4(2^3)$ design, as given by his standard orthogonal array, is only one of two possible equivalent designs; for more complex design types, such as $L_9(3^4)$, there are even more possibilities unknown to the Taguchi methods user, which can be readily constructed by those who are trained in design of experiments the traditional way (see Box *et al*.[29] for a complete coverage of the subject, including the use of *factorial* and *fractional factorial* designs).

For the purpose of studying noise parameters, Taguchi also advocated a procedure in which one matrix comprising noise parameters is embedded in the main experiment with design parameters. These matrices are referred to, respectively, as the *outer array* (since the noise parameters are external to the physical system) and the *inner array* (since the design parameters are built into the physical system). With such designs, the total number of measurements to be made could balloon considerably, draining valuable resources (time and cost) except where experimental trials can be conducted by computer simulation.

18.5.3 Data analysis

The possible drawbacks of the experimental design procedures outlined above are not the major bones of contention concerning Taguchi's parameter design. There are serious criticisms surrounding the way in which experimental data are to be analyzed. First, as pointed out before, traditional analysis of data from experimental design is based on the fundamental assumption of constant variance in response; Taguchi explicitly regards response variance as variable, and no concern is expressed over the *distributional properties of data* used in his formal statistical significance tests. Secondly, Taguchi combines the level of response y with the observed variance of y in various formats under the general label of 'signal-to-noise ratio', and suggests ways to optimize this ratio through manipulations of design parameter values. Many have argued

that with this method, opportunity is lost to adjust the parameters to suit two distinct sets of objectives, one in response level and one in response variation. Thus it has been suggested that with reference to Figure 18.6, in terms of equation (18.1), y_{I} and y_{II} could represent y and σ^2, respectively, and an investigator could determine suitable values of x_1, x_2, x_3 to meet separate objectives in level and variation control.[11,17,18]

Third, and more important, is the way in which parameter values are matched to Taguchi's orthogonal arrays for experimentation. Since the user of standard arrays is not aware of the rationale of their construction, it is difficult for him to understand the extent of confounding, that is, the mixing up of multiple parameter interactions with single-parameter effects, in his data analysis. An unsuspected interaction between two or more parameters could grossly distort the conclusion to be drawn from the analysis. The orthogonal arrays for parameters set at three levels, a popular choice of Taguchi methods users, can yield particularly misleading results. Such problems are compounded by the availability in recent years of purportedly user-friendly Taguchi methods software packages; many such packages 'automatically' select an orthogonal array, assign parameters to the columns, spew out a barrage of statistics and graphical displays, then point to a single 'best' solution, all with minimal intervention from the user.

Fourthly, in an effort to present an easy-to-follow data analysis procedure, instead of structuring a valid mathematical model for the overall input–output relationship, a routine is advocated whereby the effects of single parameters are examined one at a time, and recommended settings picked directly according to objectives such as response maximization or minimization. Such a simplistic approach, also referred to as marginal analysis or simply 'pick the winner', has been pointed out time and again[12,13] to be apt to lead to false results: this will be illustrated later in this section. Taguchi methods advocates have always defended the routine by stressing that all recommended settings are to be tested in a final confirmation experiment. However, there is no guarantee that an improved performance obtained during the confirmation experiment is indeed the best achievable, and no satisfactory suggestion has been put forward as to what is to follow if the confirmation experiment turns out to be unsatisfactory. Indeed, the question Barbra Streisand asked ruefully in the song 'The Way We Were' would not be out of place under such circumstances: 'If we had the chance to do it all again ... would we? could we?' The answer is likely to be 'no', since an ordinary Taguchi methods user, not having the capability to decipher the complete confounding pattern of interactions hidden in the orthogonal array used for his experiment, would find it impossible to try any other orthogonal array to probe the governing cause-and-effect mechanism in the subject of study.

A further objection to marginal analysis is that the recommended solution is confined only to the very parameter settings that have been used in the experimental trials; in practice, parameter fine-tuning on a range can be readily accomplished once valid mathematical models are established.[36]

In addition, Taguchi methods users generally do not have the means to break away from the experimental region to seek new parameter settings that could further optimize the desired performance, or to refine existing settings to suit changing operating conditions. Statisticians, in contrast, have on hand established strategies such as response surface methodology and evolutionary operation for such purposes. (For an illustration of response surface methodology applications that can never be carried

out with Taguchi methods, see Tan Goh.[37]) In fact, for seeking and tracking optimal parameter values during the very last stage of any optimization study, a capability for specifying an experimental design matrix, mathematical modeling, optimization and sensitivity analysis is absolutely essential, and cookbook procedures would not suit the purpose at all. There have been a host of other objections and criticisms on more theoretical grounds leveled by statisticians at Taguchi's gamut of design and analysis routines,[12,15,20,38–43] which will not be elaborated here. These, however, may appear to be of somewhat remote concern from the point of view of certain quality professionals who have come to appreciate a number of nontraditional applications that Taguchi has developed and promoted for engineers; some such applications are illustrated below.

18.5.4 Special applications

Taguchi methods stress a number of practical objectives beyond those associated with the usual SQC or experimental design studies. These objectives are often related to product or process development and design. Chief among them is the emphasis on reducing variability in the performance of a product or process; this, as already pointed out in the earlier discussion on problem formulation, is handled by using experimental design to address a changing, rather than constant, value of σ^2 of the response. When input parameters are optimized for minimum σ^2, process capability (as expressed by indices such as C_p and C_{pk}) is maximized for a given operating technology, that is, without additional capital investment.

Secondly, by introducing simulated noise parameters in experimental design, it is possible to introduce adverse manufacturing conditions and environmental stress into the study, thereby generating insight into product and process reliability in addition to quality which normally represents only stress-free and as-made characteristics. A typical performance stabilization strategy is illustrated in Figure 18.7, where it is shown that the performance index y is affected by the level of a design parameter x_d (e.g. material thickness) and that of a simulated environmental noise parameter x_e (e.g. temperature); x_d^- and x_d^+ are low and high values, respectively, of x_d used in a given experimental study, and x_e^- and x_e^+ are simulated low and high values, respectively, of x_e. It is seen that if, as a result of the experimental design study, the design

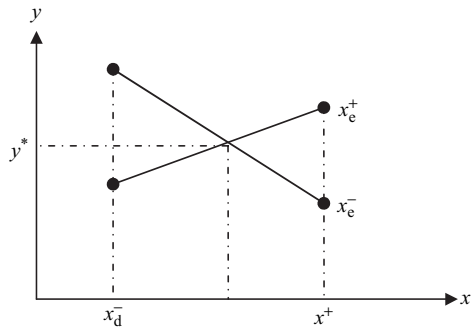


Figure 18.7 Parameter setting for robustness against noise.

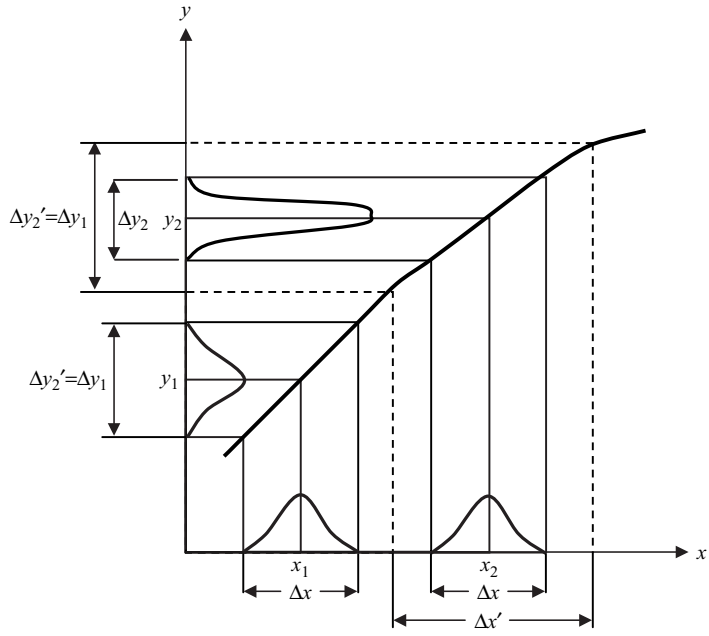


Figure 18.8 Effect of parameter x on response y.

parameter–noise relationship is revealed to be as shown in this figure and the value of x_d is then fixed at x_d^*, the value of y will from now on stay at y^* regardless of the environmental condition reflected by the actual (and in real life uncontrollable) value of x_e. Thus 'robustness' in performance is achieved without the need to invest in costly new technology or better noise-resistance materials.

Another important consideration for engineers in design and operation studies is direct cost. Apart from the loss function concept, Taguchi methods consider cost an integral part of a parameter design study. A strategy frequently cited is based on the exploitation of a nonlinear functional relationship between a manipulable parameter and response. For example, after a designed experiment has been run, the effect of parameter x on response y is isolated and can be shown graphically as in Figure 18.8. Suppose the nominal design value is originally x_1, for which the corresponding response is y_1. Because x_1 is subject to uncontrollable variation during product manufacture (owing to component-to-component variation) or usage (caused by environmental stress or deterioration), y_1 is also variable, and the amounts of variation for x and y are Δx and Δy, respectively. It can be seen that if the nominal value of x is set at x_2 instead, then for the same Δx, the variation in y is substantially reduced to Δy_2. Thus x_2 is greatly preferred to x_1 in the design, as it is relatively easy to bring the nominal value of y back from y_2 to y_1 subsequently through the use of a linear relationship between y and another design parameter. Even if a variation of Δy_1 is acceptable when compared with a given performance specification $\Delta y_2'$ ($\Delta y_2' \leq \Delta y_1$), x should not be set at x_1: as an illustration, if $\Delta y_2'$ is equal to Δy_1, then with x set at x_2, it can be seen from the figure that the hardware needed (e.g. the temperature

or speed controller used for a process, or the resistors and capacitors in a circuit) can be of a lower grade and hence lower cost, since the actual value of x can now be subject to a variation as large as $\Delta x'$ around x_2. (It could also mean that the product containing this component can be subject to more severe environment stress with no loss of performance.) In this way the design is optimized either for the most robust performance for a given cost, or for the least cost with a given acceptable level of robustness (or variability).

Overall, the problem formulation, experimental design, data analysis, and special applications of Taguchi methods constitute a package that can fit the major objectives of what has been termed robust design or quality engineering. In fact there is now a proliferation of labels applied to studies involving the application of design of experiments, for engineering objectives: Taguchi methods, off-line quality control, parameter design, quality engineering, robust design, classical design of experiments, and even robustization;[44] depending on the user, each label may or may not include a particular idea or procedure. (As an illustration of the diversity of usage of terms, it is noted that although 'classical design of experiments' is commonly taken to mean designs after Fisher and Box, there are also some who have used it in formal writing to label the ineffective 'one variable at a time' designs.[45])

18.5.5 Illustrative examples

Some simple examples will illustrate how the popular version of Taguchi's parameter design procedure, as often found in commercial courses and computer software packages, can easily be made appealing to the uninitiated. The frailty of the 'power' of parameter design, however, can be just as readily demonstrated; unfortunately that is rarely seen in such courses and packages.

Suppose the black box P represents a bonding mechanism with response y, the yield of manufactured parts with acceptable bond strength, and x_1, x_2, x_3 are design parameters denoting bonding material hold time, and contact speed, respectively. The objective of the study is to search for parameter settings to maximize the response. According to Taguchi's standard orthogonal array $L_4(2^3)$, an experiment can be carried out as shown in Table 18.2, where the choice of material x_1 is between L and H, and x_2 and x_3 are tried out at the extreme values L (for low) and H (for high) of their respective acceptable operating ranges.

With experimental results as shown in last column of Table 18.2, an analysis can be made as follows. The average of all y-values is 70 %. When x_1 is at L, the average y

Table 18.2 Illustration of a typical 'parameter design' routine.

Material x_1	Time x_2	Speed x_3	Yield $y\%$
L	L	L	72
L	H	H	82
H	L	H	56
H	H	L	70

is 77 %, 7 percent above the overall performance; when it is at H, y is 63 %, 7 percent below average. Therefore if one is to maximize y, x_1 should be set at L to extract an advantage of 7 percent. A similar reasoning for x_2 and x_3 suggests that they should be set at H and L for potential gains of 6 percent and 1 percent, respectively. Thus with this experiment, one is able to:

1. *isolate* and *quantify* the effect of each of the parameters material, time and speed on bond strength;
2. *rank* the relative importance of these parameters (material and time are much more critical than speed; there is no need to waste effort or money on speed control within the experimental range);
3. *recommend* parameter settings to maximize the yield (material L be used, time at the high end, and speed at the low end);
4. *predict* the result to be expected when the recommendations are adopted: the predicted response is 70 %, elevated by 7 percent, 6 percent, and 1 percent through 'optimized' material, time, and speed settings, respectively, giving a total of 84 %.

The striking points about this study are as follows:

1. Instead of the at least tens of experimental values one would normally expect from an investigation of three independent variables, only as few as four are needed to draw some very useful conclusions, hence the *economy of effort* is impressive.
2. Traditional investigations tend to involve a large number of trial conditions, with the eventual recommended solution based on the condition associated with the best observed result: in the present scheme, Taguchi's simple routine has revealed that a combination (L–H–L) that has never been tried out is actually superior to any of the four combinations already experimented with, thus expanding the horizon in the search for a better performance from a black box.
3. The result to be given by the recommended settings can be *expected ahead* – in this particular case it is 84 %, better than the best performance (82 %) ever actually obtained by the investigators. This makes it very convenient for shop-floor personnel to submit their recommendation for approval, since the expected benefit (84 % versus 82 %) can *immediately* be weighed against other practical considerations (for example, material L may be more expensive than material H; it may incur supply, storage, or environment problems; a lower speed means a longer product cycle time, and so on).

It is little wonder that such a 'cheap and good' and, above all, 'easy to understand' routine was able to take America by storm when first marketed. Space does not permit a comprehensive critique of the routine here, but another simple example would immediately demonstrate why enthusiasm for and confidence in at least this portion of Taguchi methods would be rather ill placed.

Suppose a similar study is conducted with two design parameters x_1 and x_2. Experiments A and B shown in Table 18.3 are both straightforward in terms of design and analysis. With reference to experiment A, to maximize y, the same routine just used will give H and L as the desirable settings for x_1 and x_2, respectively. In this case since the H–L combination has already been tried out as part of the experiment, the

Table 18.3 Two similar experiments in 'parameter design'.

x_1	x_2	y (%)
Experiment A		
L	L	37
L	H	34
H	L	50
H	H	47
Experiment B		
L	L	40
L	H	32
H	L	48
H	H	50

corresponding response of 50 % can be seen to be indeed the best that one can get out of the system.

So far so good; what about experiment B? It can easily be verified that in this case, the recommended combination is again H–L, which also has been actually tried out, but 48 % is not the best ever obtained! Since a 'confirmation experiment' as recommended by Taguchi could not possibly shed further light on this paradox (one would simply get another 48 %, assuming no repeatability error), the outcome immediately poses a dilemma:

1. Should one now go back to the good old practice of choosing the combination associated with the best observed performance, i.e. H–H, and abandon the Taguchi analysis altogether?
2. Having had one's confidence in Taguchi analysis shaken, how should one feel about the results and the very seductive interpretations of the previous experiment (Table 18.2)?

The above is indeed good material for a homework assignment for university students, but a far better purpose would be served if it were made one of the 'must ask' items by people rushing to attend one of those 'application-oriented', 'look, Ma: no mathematics!' Taguchi methods short courses.

18.6 CULTURAL INSIGHTS

It is remarkable that whereas some statisticians and academics have in recent years begun to acknowledge the sensibility of certain of Taguchi's engineering ideas,[15,17,19] the Taguchi camp has shown few signs of compromise, steadfastly vouching for the unchallengeable effectiveness of the Taguchi package of procedures in its original form. Thus it would be useful to turn to some nontechnical aspects of Taguchi methods at this point, for a better understanding of the Taguchi phenomenon, an unquestionably significant part of the history of industrial quality assurance.

18.6.1 Originality of statistical ideas

Many adverse comments have been made as regards the originality, or the absence of it, of the experimental design principles at the core of Taguchi methods.[15] To set the record straight, there is no evidence to suggest that Genichi Taguchi created the label 'Taguchi methods' for himself. In Japan, Taguchi's contributions to applied design of experiments are well known, but the subject is generally recognized by its generic name, not by 'Taguchi methods'. As for fundamental contributions, it is fair to state that the Japanese have not been known as great originators of technology anyway; most patents in technology breakthroughs in the past have their origins in the USA and Europe.

However, in modern times the Japanese have been exceedingly successful in methodically adapting, transforming, packaging and marketing acquired technology. This has been amply demonstrated by their success with consumer goods, for example electronic products and automobiles. In the field of quality control, the West has already witnessed the acquisition, simplification, and promotion of statistical tools in Japan,[46] and the subsequent exporting and marketing of such tools back to their very places of origin. The emergence of Taguchi methods is essentially only the latest exemplification of this familiar cycle.

Although it is academically correct to point out that Taguchi's orthogonal arrays and fractional factorial experimentation techniques have existed for decades, and ideas such as effects of inaccuracies in parameter setting (error transmission) are not new and could be more rigorously studied,[47,48] it may be worthwhile to ponder on the practical ramifications of Taguchi methods: stripping experimental design methodologies of their mathematical cloaks, casting them in standard formats, and putting them within the reach of rank-and-file technical personnel striving to meet quality and reliability objectives in design and manufacturing. Experimental design can now become an everyday occurrence, not something to be seen on special occasions. This by itself is a novel strategy for pervasiveness in the realization of the power of experimental design in industry, in much the same way as quality should ultimately be pervasive in society.

18.6.2 Cookbook instructions for users

One of the well-known Japanese cultural traits is the propensity for uniformity and doing things by standard examples. Foreign visitors to Japan are invariably struck by the omnipresence of realistic wax models (*mihon*) of food dishes exhibited in showcases outside restaurants. At most establishments only what is on display is offered, and a customer is expected to order from the standard fare with no possibility of variations – and within minutes he gets exactly what the model shows. Virtually all eating places also offer set meals (*teishoku*), saving the customer the trouble of making decisions and creating combinations of orders, and the cooks some preparation time as well. It would not be trivial at all to draw the parallel between the mind-sets governing the way restaurant dishes are offered, chosen and consumed, and the way Taguchi's parameter design techniques are presented (through standard orthogonal arrays), adopted (by means of a collection of artfully crafted linear graphs), and executed (step-by-step marginal analysis, graphical presentations, and analysis of variance).

It should be realized that, being highly literate as a nation, the Japanese would find looking up published tables and graphics a common daily experience. They are used to elaborate but published procedures, rules and regulations. In much the same way, a parameter design engineer at work can simply turn to the collection of standard orthogonal arrays, linear graphs, and interaction tables; he does not have to question the background of array structures, for example, and is never required to construct individual fractional factorial matrices from scratch. He would not be aware of the fact that orthogonal arrays offered by Taguchi in his books are not the only ones available; there are in fact many more possibilities in the design of fractional factorials, depending on the choice of parameter arrangements and level of resolution of parameter effects – a foot soldier in the Taguchi camp would be oblivious to all such possibilities for creative design and applications. The procedures for data analysis are even more well defined: analysis of variance is but a series of mechanical calculations; mathematical modeling concepts need not be involved. The final confirmation experiment winds up the project, in much the same way as a loud burp – commonplace in Japanese custom – signifies the completion of a satisfactory meal. It may be argued, of course, that the advantage of conducting and analyzing experiments via recipes is that the experiment can be quickly designed, performed and analysed, and actions recommended – all accomplished with nothing more than elementary algebraic skills.

18.6.3 Aversion to variations, conflicts and uncertainties

There is a host of other Japanese cultural habits that reflect the inclination toward uniformity, harmony, and predictability. It has already been pointed out that while American 'gurus' talk about pleasing the next customer (and not any other), Taguchi thinks of helping society. Unlike Westerners, Japanese businessmen have clearly defined and observed dress codes and protocols. A Westerner opens the present that he has just received immediately, whereas a Japanese does not rush to do so. Americans find it difficult to appreciate the merits of packaged group tours, but such tours are the norm for Japanese vacationers. American companies hire and fire as necessary, Japanese companies emphasize employee loyalty instead. Western statisticians think nothing of switching from maximization to minimization in optimization studies, but Taguchi deemed it desirable to concoct a variety of 'signal-to-noise ratios' that are *all* to be maximized (by building in a minus sign or using reciprocals whenever minimization is required). Indeed, with a signal-to-noise ratio, the potential conflicts (in terms of recommended parameter settings) between the requirements for optimizing the response level and minimizing response variation will be conveniently hidden or even forgotten, since only compromise solutions are given by the analysis of such ratios.

Taguchi's linear graphs and confounding tables reveal only two-factor interactions; potential three-factor (or higher-order) interactions are a non-subject even though sometimes they can be significant. It is well known that while statisticians are concerned about the potential difficulties that could be caused by confounding of interaction effects in a fractional factorial experiment, Taguchi does not dwell on confounding problems, and avoids conflicts or uncertainties in the interpretation of experimental results by advising engineers not to consider interaction effects at all unless they are already known to exist. There is in fact a message that interactions should be 'avoided', as if interactions, which arise from natural mechanisms and phenomena, can be

eliminated by some manner of self-blindfolding (somewhat in the spirit of *kamikaze*) – witness the following statement in a recent journal paper:

> *Interaction was the most confusing element of the Taguchi methodology to the team members. Some of the assumptions and decisions made were based purely on the team members' instinct, rather than scientific foundations. I found that this point of the project was the most divisive to the team. In subsequent Taguchi's experiments, I have favoured the use of non-interacting orthogonal array tables such as* L_{12}, L_{18} *and* L_{36}.[49]

Simply by choosing Taguchi arrays that are usable only under the condition of absence of interaction and useless otherwise, one can get rid of troublesome natural interactions in a stroke! That is a most remarkable revelation of the heights of faith to be found among Taguchi followers.

18.6.4 Indoctrination and personal cult

The generation of Taguchi faithful first emerged in the early 1980s among nonstatisticians. Before their first serious treatment in an academic context,[8] Taguchi methods as depicted in professional magazines[21–24] tended to be studded with staggering claims sometimes shrouded in tantalizing and almost mystical depictions – a mix of oriental inscrutability in experimentation schemes and the awesome reality of Japanese competitiveness in product quality. In addition, the honorifics 'Doctor' and 'Professor' were indispensable whenever Genichi Taguchi's name was mentioned, a practice that is sustained even today in many books and journals. Intentionally or otherwise, such a build-up of reverence for an oriental master has led to a cult of sorts for a larger-than-life image. In fact, it is notable that a feature of Taguchi's works is the use of Socratic dialogs between a master and a disciple to clarify ideas and techniques; another is the repeated use of case studies in place of general formulations to present a subject, very much in the ancient tradition of teaching by parables.

Meanwhile, the American Suppliers Institute was set up in Michigan for the primary purpose of training automobile suppliers. It functioned as the US Center for Taguchi Methods for years, running a stream of short training courses on the subject. It has also been running annual Taguchi symposia – with published proceedings[50] – at which one speaker after another would bear witness to the power of Taguchi methods in improving quality and reliability, vouch for the role of Taguchi methods in securing huge savings for companies, and tell countless other impressive success stories. (Understandably no accounts of failure ever got a chance to reach the symposia.) The fervor in fact even spread to the opposite shores of the Atlantic, where papers on Taguchi methods in the ASQC's *Quality Progress*[9,51] were reprinted in *Quality Assurance*, the journal of the UK Institute of Quality Assurance; in a typically British manner a Taguchi Club was duly formed, and Taguchi courses and conferences soon started to spread across the Continent.[52,53] It is worth noting that no corresponding movements took hold in the Orient, although Taguchi's works had been well studied and cited in mainland China, Taiwan and India years before the Americans became aware of them.[54,55]

The Taguchi camp, after more than a decade of self-reinforcement of its beliefs, is on the whole quite unperturbed by criticisms of weaknesses in the methods' theoretical

foundations and likely pitfalls awaiting their users. Maintaining a studied silence on concerns such as the potential problems with the $L_9(3^4)$ design (not to mention the even more theoretical issues), the camp has found (and continues to find) it difficult to bring itself to openly conceding any shortcomings in the Master's teaching.

18.6.5 Backlash and alternatives

It may be noted that while there may be excessive passion among Taguchi's followers, some criticisms against Taguchi methods have been no less pointed. Thus 'accumulation analysis is not a useful technique and to teach it to engineers would be a serious mistake'.[41] And 'maybe in other parts of the world, you can sell engineers simplistic solutions as profound enlightenment. But around here, we don't just buy something on faith ... Mr. Taguchi, pack up your "methods" and get out of my lab. DO NOT TALK to any of my engineers or technicians. Do not suggest any other ways to be more "efficient" or "optimized". Just get out'.[56] The latter quote did not come from a statistician, but an engineer disappointed with what Taguchi methods purport to offer to improve engineering design. More questions have also been raised in engineering circles lately.[57,58] Actually, in terms of 'market share' today, Taguchi methods appear to have an edge over the other 'brands' of experimental design (Box, Shainin, etc.) outside academia. Although the use of statistical designs in a special way without acknowledging their sources may not be as serious as industrial espionage, which the Japanese have from time to time been accused of, American industry is now being warned against being unwittingly done in by the Japanese (or Japanese methods) on the engineering front, not just in trade. In fact, a curious love–hate feeling, not unlike what takes place in a Toyota or Honda showroom, has been apparent during many conference or workshop discussions on Taguchi methods in the USA.

In any case, as is now generally agreed, Taguchi methods on balance do have the potential to bring about first-cut improvements in industrial applications. However, owing to their theoretical imperfections and other limitations, success cannot be assured in every instance, much like an angler who can go to the water conveniently, but cannot be sure of a catch every time a line is thrown. (This point was amply demonstrated earlier on through the of experiment A and B in Section 18.5.5.) Applied statisticians would rather get the basic philosophy laid out and understood,[14] and go for methodologies that are on firm theoretical grounds which can lead users to the best possible results. Exemplified by response surface methodology,[30] the aim is the best solution, not marginal improvements, hence their strategy is more akin to mountaineering. Herein lies the contrast in choices: should one opt for the routines of an oriental set piece for fishing, hoping for the best, or should one acquire all the necessary tools to meticulously execute an ambitious, high-reward mountain climbing expedition, exercising individual insights and judgments along the way?

18.7 TRAINING AND LEARNING

The above metaphorical alternatives have emerged from debates between Taguchi advocates and skeptics, which are never short of colorful claims and counterclaims. The Taguchi controversy is damaging when it comes to introducing statistical design

of experiments in industry: practitioners, although not particularly attracted by abstract arguments brandished by statisticians, will also be wary of embracing Taguchi methods when they see counterexamples that demonstrate the problems of Taguchi recipes, as well as open and direct condemnations of the Taguchi approach. Thus the controversy, the scale of which has never been seen before in the quality profession, could prove counterproductive if managers and engineers are to be perplexed into inaction when faced with suggestions to get into experimental design, the source of the most powerful techniques for fundamental quality improvements. Therefore it would be useful to consider strategies by which industrial personnel may possibly be led out of the woods and inducted into the community of experimental design beneficiaries.

18.7.1 Motivation

As already pointed out, experimental design was historically developed for the understanding of complex natural phenomena via empirical means. Taguchi extended its applications from a mere understanding of existing natural phenomena to the creation process of human-made physical mechanisms. Taguchi methods offer their users a structured approach to meeting performance objectives set for a manufacturing process or manufactured product in the face of variations induced by raw materials, basic components, and the environment, expressing the degree of success of the effort in a language that management can understand, that is, in monetary terms via a special quality definition and the loss function concept.

To industrial personnel it may be prudent not to speak of experimental design *per se* since it tends to invite subconscious resistance by those who have a mental block against probability and statistics. Instead, robust design or quality engineering motives should be presented to them as a capability for performance enhancement, in terms of problem solving or design optimization, that does not require a search for new technology or additional capital investment. For organizations that are not unfamiliar with mathematical ideas, however, the subject of experimental design can be quite logically introduced as further tools to be acquired for fundamental gains and breakthroughs after SQC or statistical process control has been stretched to the limits of its usefulness. In this respect the summary statements shown in Table 18.4, in addition to Table 18.1, would be useful for motivation as well as appreciation purposes.

18.7.2 Applications

Since statistical quality engineering is a relatively new subject, most people in industry have to learn it from short training courses designed for working personnel. Suggestions for effective design and implementation of training programs have been offered elsewhere.[59] An important point is that to make an application-oriented course useful, course contents should be designed in accordance with the 80–20 principle, reflecting the fact that as much as 80 % of all problems encountered in practice are likely to be adequately handled by 20 % of all available techniques. It would therefore be unproductive to pursue a complete range of topics, a practice more for academic courses in colleges and universities than industrial training programs. In this way some of the

Table 18.4 Design of experiments vs. statistical process control.

Statistical process control	Statistical experimental design
1. Used for 'on-line' quality control	1. Used for 'off-line' quality engineering
2. Deals mainly with existing processes	2. Can be applied at process design and development stage
3. Meant for routine application	3. Has a problem-solving dimension
4. Aims to maintain status quo .	4. Seeks improvements and best operating states
5. No new operational targets	5. Motivated by specific needs and seeks new results
6. Nonintervention of physical system	6. Purposeful probing of physical system
7. Based on passive observation of system output	7. Depends on active manipulation of system input–output linkages
8. Monitors known key parameters	8. Identifies key parameters
9. No forward planning element	9. Attempts to foresee and prevent problems
10. Waits for problems to happen	10. Identifies sources of problems and seeks their elimination
11. No obvious sense of urgency	11. Efficiency is important
12. Carried out continuously	12. Carried out project by project

more controversial Taguchi procedures need not be introduced to nonspecialists in first courses and, by the same token, less down-to-earth statistical arguments can be reserved for those who have the inclination to probe deeper into the subject.

In place of an emphasis on completeness of coverage, efficiency of information extraction and utilization should be highlighted as the prime concern. As is well known, Taguchi methods users have to base their experimental designs on standard tables that offer very little latitude in their adoption; thus one has to be careful not to let design of experiments be reduced to an exercise in which a problem is made to fit a standard orthogonal array. It is not unusual to see Taguchi case studies involving multilevel designs right from the outset. In practice, however, fewer experimental runs with two-level fractional factorials should be attempted first as (i) not all parameters will turn out to be significant (again the 80–20 principle), so no purpose is served by according to every one of them a multilevel treatment, thereby increasing the size of the experiment; (ii) most changes in response are approximately linear with respect to parameter values that vary within a realistic experimental range; (iii) even if nonlinearities exist, they do not bias the estimates of main and interaction effects; (iv) it is in fact possible to institute simple tests for the existence of significant nonlinearities and, if they do exist, (v) it is a straightforward matter to augment a two-level design by additional experimental runs to constitute a higher-level design.[12,29,30] Hence it is unproductive to plunge into grandiose designs, complete with inner and outer arrays, and bet on the final confirmation experiment to justify the entire effort, when fruitful results can be obtained through a judiciously executed series of small, sequential experiments.

Another important feature of an effective training program would be the elucidation, best done by case studies, of the roles of different tools at different points in the learning curve associated with the study of a particular physical subject. Taguchi

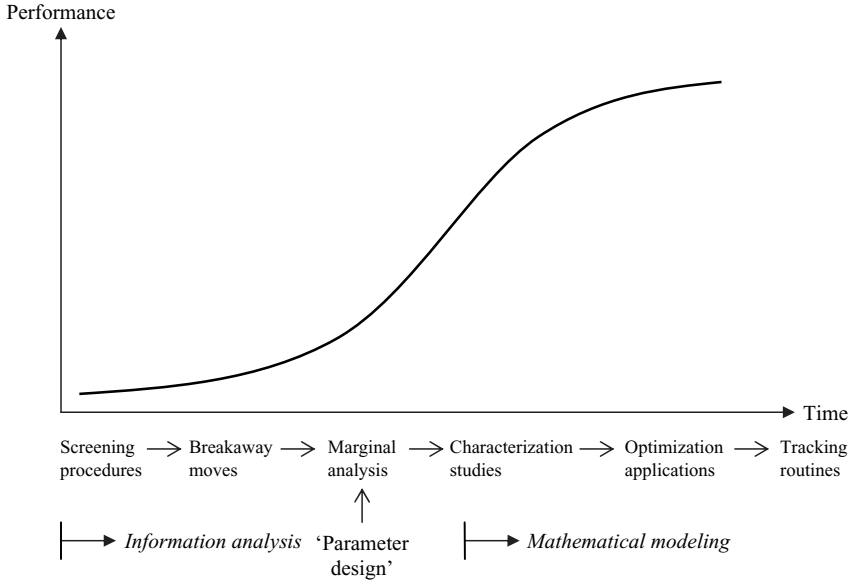


Figure 18.9 Choice of statistical tools along the learning curve.

methods, as already pointed out, offer a prescribed set of techniques, whereas statisticians represented by Box encourage *à la carte* selections used according to the progress of investigation: screening, characterization or optimization. Thus the actual design of an experiment is shaped by the existing knowledge about the subject of study, and the size of experiment is defined according to the expectation on the quality of the resulting information (e.g. level of resolution of parameter effects). In this respect the Box approach is much more versatile and flexible, as it offers opportunities for quick screening of a large number of parameters, followed by detailed mathematical modeling of the critical ones that remain or, in some cases, a quick breakaway from current ranges of parameter values for better results, culminating in a final fine-tuning of parameter settings for optimized performance. Indeed on-line optimality seeking (by response surface methodology) and tracking (by evolutionary operation) are strategies that are sorely absent from Taguchi methods; no statistical quality engineering training or application would be complete without their inclusion. Figure 18.9 shows the sequence of major tools to be applied as performance progresses along the learning curve.

18.7.3 Directions

In discussing the technicalities of statistical quality engineering with management and operations personnel, measures should be taken to ensure that they do not fail to see the forest for the trees. Hence, it would be useful to explain the historical background of the various schools of thought that exist in the profession today. An awareness should be brought about of the richness of experimental design literature, scope of applications, and practical conditions to be encountered in practice,[60,61] and

Table 18.5 Approaches and concerns in experimental design applications.

Mainstream statistics	Taguchi methods
How experiments can be conducted outside the laboratory for the study of an operating system	How experiments can be tied to engineering design and cost optimization
How to understand the true nature of the object of study by means of inductive reasoning	How to translate data into engineering conclusions and hence specific actions
How to secure valid theoretical foundations for the resulting conclusions and decisions	How to reduce the entire methodology into practical procedures for nonstatisticians
How to represent significant cause-and-effect relationships in a product or process by mathematical models for performance optimization	How to insulate a product or process from both present and future external causes of performance deterioration
How to obtain the optimal – not suboptimal	How to obtain a working solution

the strategic ideas behind each modern practice such as Taguchi's should well explained and not drowned in a plethora of technical details and analysis procedures. Bearing in mind the difference in perspectives between statisticians and engineers,[62] a discussion along the lines of Table 18.5 would serve a useful purpose.

Admittedly, at one point or other, the question of 'brand selection' will surface, but one should not feel obliged to declare one's affinity or loyalty to any, inasmuch as there is considerable common ground among the alternatives. For example, techniques propounded by the Taguchi and Box schools are by no means mutually exclusive, and a clearer understanding of both can be gained by means of a presentation such as Figure 18.10. Indeed the figure reflects but a facet of two overlapping yet different approaches, one stressing robustness in the engineering sense, another robustness with its traditional meaning in statistical inference. Polarization of approaches taken by engineering-based Taguchi methods (TM) and statistics-oriented methodologies typified by Box, Hunter and Hunter (BH^2) has created different routes to quality and reliability improvement. To return to the fishing and mountaineering analogy, there

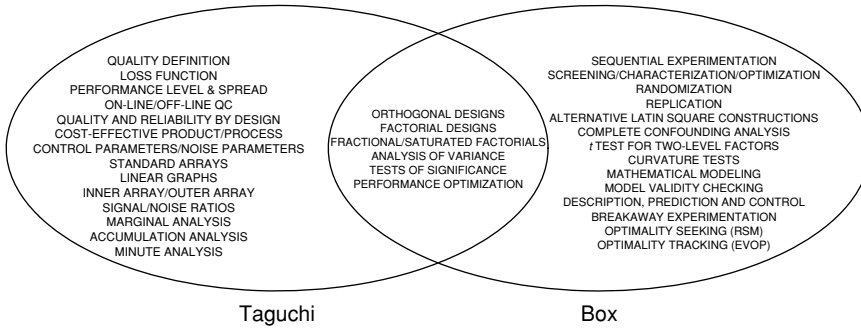


Figure 18.10 Overlapping and distinct areas of Taguchi methods and mainstream statisticians' methodologies.

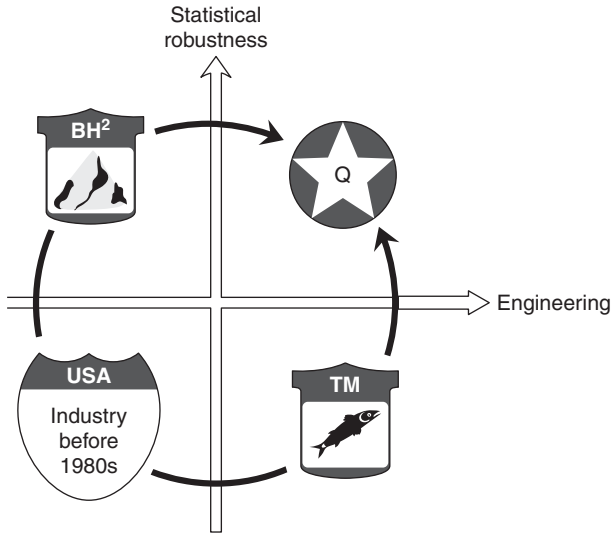


Figure 18.11 Routes to quality via statistical and engineering robustness.

may be more people enjoying fishing than mountaineering, but the discerning and committed on either route will not fail to be duly rewarded in the end. Many American companies in particular, notwithstanding the intensity of debates on Taguchi methods found between the covers of academic journals, have already decided for themselves on the specific route to take (witness their participation at the Taguchi symposia, for example); hopefully, in not too distant a future, all will recognize which techniques are reliable and which are not, and begin to attain a maturity in technique selection that will ensure both engineering and statistical robustness, as suggested figuratively in Figure 18.11.

18.8 CONCLUDING REMARKS

Taguchi methods are important because never before in the history of industrial quality assurance has a set of ideas and methodologies formulated by a single person created so much interest and so much controversy in so short a time among so many people in so many spheres – from the shop-floor to the corporate boardroom and from hotel seminar rooms to the university lecture theaters. Techniques labeled as Taguchi methods have in fact long been known in the Orient by their generic names; the sudden surge of interest in Taguchi's teaching in the USA in the early 1980s was not without its historical and cultural background, the most immediate being the impact on the international competitiveness of US industries by Japanese achievements in product quality. Quality improvement methods carrying a Japanese name during this period would without exception quickly attract attention in the USA, but there are also some other reasons why interest in Taguchi methods soon turned into unprecedented zeal and faith.

Basically, Taguchi methods address all the major quality improvement expectations, shaping a strategy that promises to integrate the management, technology, and information capabilities in an organization. They throw up an unambiguous bottom line in a language that everyone in the profession understands: on-target, least-variation and cost-effective performance throughout the life of a product or process in the real world. An offering like this, designed with a mix of oriental idealism and pragmatism, presented in Japanese consumer-oriented packaging and promoted with American marketing prowess,[63] was nothing short of formidable in a country looking for a way out of the quality crisis. With parameter design at its core, Taguchi methods draw heavily from techniques of statistical design and analysis in a manner that facilitates immediate adoption by nonstatisticians. In so doing, the American quality profession was offered what may be termed 'statistical engineering', of which a practice-oriented approach has won the approval of those still uncomfortable with the 'engineering statistics' available from statisticians.

As it turned out, Taguchi methods' demonstration effect grew entirely out of proportion to their technical superiority. Controversies ensued before long over a number of ways in which statistical techniques are deployed in Taguchi methods. Disagreements have led to new studies and research in academia that could eventually increase the effectiveness of practical techniques, for example the use of graphical techniques for experimental design[64–66] and the exploitation of response surface methodology for variance reduction.[67,68] However, debates have also in the meantime led to bewilderment in the quality profession and hinded the application of useful experimental design techniques.

For the benefit of practitioners, this chapter has highlighted the nature of Taguchi methods as well as a number of issues frequently raised about them. The subject has been interpreted from several angles in the hope that understanding is the most logical approach to resolving differences. Managers and engineers in particular have a need to get out of the woods; they cannot afford to wait for the dust to settle among the advocates and critics of Taguchi methods. There has been, however, some evidence of abatement of excesses. One example is the recent 'balanced' assessment of Genichi Taguchi (score: 10 out of 10).[19] On the other side of the Atlantic, the UK Taguchi Club has now been renamed the Quality Methods Association in recognition of the fact that no single guru should claim monopoly over the faith of quality professionals. The new robust design software from AT&T Bell Laboratories has intentionally discarded the flawed analysis of variance routine in Taguchi-style experiments. These are encouraging signs suggesting that reason and objectivity will indeed win the day. When that happens, nothing but good will be brought to management, engineering, and operations personnel in industry and, by extension, society as a whole – which of course means that the mission to cut loss to society is finally accomplished.

18.9 EPILOGUE

Since there can be no argument as to the usefulness of statistics in quality improvement efforts – despite differences in the means of its deployment, some of which have

been discussed in this chapter – it would not be inappropriate to present here a little 'alphabet story' which this author has created, for effect, for his management seminars:

At the	**N**ow:
Beginning:	**O**verall
Countless	**P**roductivity,
Defects,	**Q**uality and
Endless	**R**eliability via
Faulty	**S**tatistical
Goods,	**T**echniques
Huge	**U**sed
Inventory of	**V**ery
Junk,	**W**idely to
Killing	**X**perience
Losses to	**Y**ear-round
Management.	**Z**ero defects.

REFERENCES

1. Taguchi, G. (1987) *System of Experimental Design*, Vols. 'one' and 2. White Plains, NY: UNIPUB/Kraus. New York.
2. Taguchi G. (1986) *Introduction to Quality Engineering*. Tokyo: Asian Productivity Organization.
3. Taguchi, G. and Wu, Y. (1985) *Introduction to Off-Line Quality Control.* Nagoya: Central Japan Quality Control Association.
4. Barker, T.B. (1985) *Quality by Experimental Design*. New York: Marcel Dekker.
5. Ross, P.J. (1988) *Taguchi Techniques for Quality Engineering*. New York: McGraw-Hill.
6. Phadke, M.S. (1989) *Quality Engineering Using Robust Design*. Englewood Cliffs, N.J: Prentice Hall.
7. Logothetis, N. and Wynn, H.P. (1989) *Quality through Design: Experimental Design, Off-Line Quality Control and Taguchi's Contributions*. Oxford: Clarendon Press.
8. Kacker, R.N. (1985) Off-line quality control, parameter design, and the Taguchi method (with discussion). *Journal of Quality Technology*, 17, 176–209.
9. Kacker, R.N. (1986) Taguchi's quality philosophy: analysis and commentary. *Quality Progress*, 19, 21–29. Also *Quality Assurance*, 13, 65–71 (1987).
10. Hunter, J.S. (1985) Statistical design applied to product design. *Journal of Quality Technology*, 17, 210–221.
11. Hunter, J.S. (1987) Signal-to-noise ratio debated (letter to the Editor). *Quality Progress*, 20, 7–9.
12. Hunter, J.S. (1989) Let's all beware the latin square. *Quality Engineering*, one, 453–465.
13. Ryan, T.P. (1988) Taguchi's approach to experimental design: some concerns. *Quality Progress*, 21, 34–36.
14. Box, G.E.P. and Bisgaard, S. (1987) The scientific context of quality improvement. *Quality Progress*, 20, 54–61.
15. Box, G.E.P., Bisgaard, S. and Fung, C.A. (1988) An explanation and critique of Taguchi's contributions to quality engineering. *Quality and Reliability Engineering International*, 4, 123--131.
16. Taylor, G.A.R. *et al*. (1988) Discussion on Taguchi. *Quality Assurance*, 14, 36–38.
17. Pignatiello, J.J. (1988) An overview of the strategy and tactics of Taguchi. *IIE Transactions*, 20, 247–254.

18. Tribus, M. and Szonyi, G. (1989) An alternative view of the Taguchi approach. *Quality Progress*, 22, 46–52.
19. Pignatiello, J.J. and Ramberg, J.S. (1991) Top ten triumphs and tragedies of Genichi Taguchi. *Quality Engineering*, 4, 211–225.
20. Nair, V.N. (ed.) (1992) Taguchi's parameter design: a panel discussion. *Technometrics*, 34, 127–161.
21. Sullivan, L. P. (1984) Reducing variability: a new approach to quality. *Quality Progress*, 17(7), 15–21.
22. Burgam, P.M. (1985) Design of experiments – the Taguchi way. *Manufacturing Engineering*, 94, 44–47.
23. Ryan, N. E. (1987) 'Tapping into Taguchi'. *Manufacturing Engineering*, May, 3–46.
24. Sullivan, L.P. (1987) The power of Taguchi methods. *Quality Progress*, 20, 76–79.
25. Grant, E.L. and Leavenworth, R.S. (1988) *Statistical Quality Control*, 6th edn. New York: McGraw-Hill.
26. Fisher, R.A. (1960) *The Design of Experiments*, 7th edn. Edinburgh: Oliver and Boyd.
27. Bisgaard, S. (1992) Industrial use of statistically designed experiments: case study references and some historical anecdotes. *Quality Engineering*, 4, 547–562.
28. Davies, O.L. (ed.) (1954) *The Design and Analysis of Industrial Experiments*. Edinburgh: Oliver and Boyd.
29. Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978) *Statistics for Experimenters*. New York: Johan Wiley & sons, Inc.
30. Box, G.E.P. and Draper, N.R. (1987) *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons, Inc.
31. Khuri, A.I. and Cornell, I.A. (1987) *Response Surfaces: Design and Analysis*. New York: Marcel Dekker.
32. Box, G.E.P. and Draper, N.R. (1969) *Evolutionary Operation*. New York: John Wiley & Sons, Inc.
33. Daniel, C. (1976) *Applications of Statistics to Industrial Experimentation*. New York: John Wiley & Sons, Inc.
34. Diamond, W.J. (1981) *Practical Experiment Designs for Engineers and Scientists*. Belmont, CA: Wadsworth.
35. Shainin, D. and Shainin, P.D. (1990) Analysis of experiments. *Transactions of Annual Quality Congress*, ASQC, Milwaukee, Wisconsin, pp. 1071–1077.
36. Goh, T.N. (1989) An efficient empirical approach to process improvement. *International Journal of Quality and Reliability Management*, 6, 7–18.
37. Tan, E. and Goh, T.N. (1990) Improved alignment of fibre-optics active devices via response surface methodology. *Quality and Reliability Engineering International*, 6, 145–151.
38. Nair, V.N. (1986) Testing in industrial experiments with ordered categorical data (with discussion). Technometrics, 28, 283–308.
39. Leon, R.V., Shoemaker, A.C. and Kacker, R.N. (1987) Performance measures independent of adjustment: an explanation and extension of Taguchi's signal to noise ratio (with discussion). *Technometrics*, 29, 253–285.
40. Box, G.E.P. (1988) Signal to noise ratios, performance criteria and transformations (with discussion). *Technometrics*, 30, 1–40.
41. Box, G.E.P. and Jones, S. (1990) An investigation of the method of accumulation analysis. *Total Quality Management*, 1, 101–113.
42. Hamada, M. and Wu, C.F.J. (1990) A critical look at accumulation analysis and related methods (with discussion). *Technometrics*, 32, 119–162.
43. Hamada, M. (1992) An explanation and criticism of minute accumulating analysis. *Journal of Quality Technology*, 24, 70–77.
44. McEwan, W., Belavendram, N. and Abou-Ali, M. (1992) Improving quality through robustisation. *Quality Forum*, 19, 56–61.
45. Dale, B.G. and Cooper, C. (1992) Quality management tools and techniques: an overview (part II). *Quality News*, 18, 229–233.
46. Ishikawa, K. (1976) *Guide to Quality Control*. Tokyo: Asian Productivity Organization.

47. Box, G.E.P. (1963) The effects of errors in the factor levels and experimental design', *Technometrics*, 5, 247–262.
48. Fung C.A. (1986) Statistical topics in off-line quality control. Ph.D. thesis, Department of Statistics, University of Wisconsin-Madison.
49. Shina, S.G. (1991) The successful use of the Taguchi method to increase manufacturing process capability. *Quality Engineering*, 3, 333–349.
50. American Supplier Institute (1984-1990) *Proceedings of ASI Symposia on Taguchi Methods*, Vols. 1–8. Dearborn, MI: American Supplier Institute.
51. Barker, T.B. (1986) Quality engineering by design: Taguchi's philosophy. *Quality Progress*, 19, 32–42. Also: *Quality Assurance*, 13, 72–80 (1987).
52. Bendell A. (ed.) (1989) *Taguchi Methods: Proceedings of the 1988 European Conference*. London: Elsevier.
53. Bendell, A., Disney, J. and Pridmore, W.A. (eds) (1989) *Taguchi Methods: Applications in World Industry*. IFS Publications, Bedfordshire, U.K.
54. Academia Sinica, Changyong Shuli Tongji Fangfa (1973) (Common Statistical Methods, with nine case studies), *Science Press*, Beijing, (in Chinese).
55. Zhengjiao Shiyan Shejifa (1975) (Design of Orthogonal Experiments), *Shanghai Science and Technology Press*, Shanghai, (in Chinese).
56. Pease, R.A. (1992) What's all this Taguchi stuff, anyhow? *Electronic Design*, 25 June, 83–84.
57. Sprow, E.E. (1992) What hath Taguchi wrought? *Manufacturing Engineering*, April, 57–60.
58. Smith, J. and Oliver, M. (1992) Taguchi: too good to be true? *Machine Design*, 8 October, 78–79.
59. Goh, T.N. (1992) An organizational approach to product quality via statistical experiment design. *International Journal of Production Economics*, 27, 167–173.
60. Steinberg, D.M. and Hunter, W.G. (1984) Experimental design: review and comment. (with discussion). *Technometrics*, 26, 71–130.
61. Hahn, G.J. (1984) Experimental design in the complex world. *Technometrics*, 26, 19–31.
62. Hoadley, A.B. and Kettenring, J.R. (1990) Communications between statisticians and engineers/physical scientists' (with commentary). *Technometrics*, 32, 243–274.
63. Taguchi, G. and Clausing, D. (1990) Robust quality. *Harvard Business Review*, January/February, 65–75.
64. Tsui, K.L. (1988) Strategies for planning experiments using orthogonal arrays and confounding tables. *Quality and Reliability Engineering International*, 4, 113–122.
65. Kacker, R.N. and Tsui, K.L. (1990) Interaction graphs: graphical aids for planning experiments. *Journal of Quality Technology*, 22, 1–14.
66. Wu, C.F.J. and Chen, Y. (1992) A graph-aided method for planning two-level experiments when certain interactions are important, *Technometrics*, 34, 162–175.
67. Vining, G.G. and Myers, R.H. (1990) Combining Taguchi and response surface philosophies: a dual response approach. *Journal of Quality Technology*, 22, 38–45.
68. Myers, R.H., Khuri, A.I., and Vining, G.G. (1992) Response surface alternatives to the Taguchi robust parameter design approach. *American Statistician*, 46, 131–139.

19

Economical Experimentation via 'Lean Design'

T. N. Goh

In industrial applications of design of experiments, practical constraints in resources such as budget, time, materials and manpower could lead to difficulties in completing an experiment even of moderate size. This problem can be addressed by an approach referred to as 'lean design', with which no more than a bare minimum number of response values are required, and the results of a lean design experiment can be improved incrementally up to a point identical to what can be obtained from a regular design. A numerical example illustrates both the principles and applications of this cost-effective approach to experimentation.

19.1 INTRODUCTION

Design of experiments as a means of understanding cause-and-effect relationships among the controllable factors of a process and response characteristics is now an established quality improvement tool in industry. In the past fifteen years, design of experiments has also been integrated with ideas of quality engineering for product and process design, first advocated through what have become popularly known as Taguchi methods.[1,2] Although there have been disagreements over the efficacy or even validity of Taguchi methods,[3–5] the debates have brought about a widespread awareness of a variety of issues related to the application of experimental design in real life; examples of design considerations are type of design (number of levels; what fraction of a factorial to use; resolution of effects), structure of design (taking

This chapter is based on the article by T. N. Goh, 'Economical experimentation via "lean design"', *Quality and Reliability Engineering International*, 12, 1996, pp. 383–388, and is reproduced by the permission of the publisher, John Wiley & Sons, Ltd

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

into account prior technical knowledge as well as possible confounding of effects in data analysis), size of experiments (basic number of observations together with replications), and execution plans (a single complex design, or several simpler designs conducted sequentially).

Against the above backdrop, this chapter presents a practical experimental design approach, referred to as *lean design*, which has economy of data collection as its prime objective and can address a number of problems arising from constrained resources in industrial experimentation. The procedures, based on the provisional use of incomplete orthogonal matrices, are shown to be useful for damage control in the event of inadvertent incomplete experimentation. The working of lean design is also illustrated by a numerical example based on a well-known case study in the seminal work on applied experimental design by Box *et al*.[6]

19.2 TWO ESTABLISHED APPROACHES

The strategies for lean design are developed with reference to some of the most notable features of design of experiments as advocated by the Taguchi school and by mainstream statisticians. For example, while a Taguchi design tends to be a one-shot experimental effort, mainstream statisticians would prefer using simpler designs conducted in a sequential manner – the 'fold-over' design used by Box *et al*.[6] being a well known example. With sequential experimentation, interim results can be obtained to provide insights into the behavior of the subject of study and help focus experimental efforts: for example, two suitably designed 2^{7-4} experiments in tandem would be preferable to one 2^{7-3} experiment comprising the same physical experimental runs, since the investigator using the 2^{7-3} design would have no inkling of the likely results of the experiment before all 16 observations are available. This principle will be brought up again in later discussions.

In Taguchi methods, an experiment is designed with the assumption that interaction effects among experimental factors can generally be ignored, except for those already known to be present by virtue of the technical knowledge on the part of the investigator. For this reason most designs are saturated fractional factorials, but the resulting confounding of effects in data analysis is, rightly or wrongly, not of particular concern – a subsequent 'confirmation experiment' will serve to guard against erroneous results.

Although the wisdom of ignoring interactions has often been questioned, Taguchi designs have attracted a considerable following, mainly because they make it possible to use experiments with fewer experimental runs. If such an approach is accepted and exploited in the thought processes of experimental design, then there are opportunities to reduce experimental runs in a given investigation to their bare minimum. The discussion below therefore covers situations where Taguchi's no-interaction assumption is not categorically rejected.

19.3 RATIONALE OF LEAN DESIGN

The basic principles, explained in terms of common notation used in the literature, are as follows. In an orthogonal experiment, the maximum number of independent

statistics, or values of factor effects, that can be computed is equal to the number of experimental runs. Technically, this is to say that the degrees of freedom associated with the computed effects equal the degrees of freedom carried by the raw data used in the analysis. Thus in an unreplicated 2^3 full factorial design, eight response values $y_1, y_2, \ldots, y_8$ are used to generate eight statistics E_0, E_1, E_2, E_3, E_{12}, E_{13}, E_{23}, and E_{123}. With the eight response values from a fractional factorial design, say of the type 2^{4-1}, eight statistics are generated to represent a collection of main and interaction effects – 16 in total – inherent in the subject under study. If all interaction effects are of interest, there will be confounding of two effects in every statistic; however, if interaction effects are non-existent, four clear main effects E_1, E_2, E_3, E_4, in addition to the mean E_0, can be obtained.

Without the interactions, since E_0, E_1, E_2, E_3, E_4 account for five degrees of freedom, the three remaining degrees of freedom are arguably not really needed. Taguchi analysts would create an error sum of squares to absorb the three degrees of freedom, but this is an expediency that does not result in something as reliable and useful as a pure error sum of squares obtainable from replicated experimental runs. Since these three redundant degrees of freedom must originate from three response values, one can conclude in turn that three out of eight such response values are, in this sense, unnecessary. Hence only five of the experimental runs, not eight, need actually be carried out. An experiment formulated along this line of thinking – based on an orthogonal design not meant to be carried out in full, thus achieving economy in the required experimental effort – is one of *lean design*.

It is clear from the above discussion that lean design can be said to be 'more Taguchi than Taguchi'. First, one is supposed to know what to ask for in an experimental data analysis – an a priori listing of main and interaction effects to be obtained from the experiment. Secondly, since all other interactions are assumed to be non-existent, confounding will not be a concern. Furthermore, as long as the total number of effects asked for is less than what the size of the design matrix would provide, one has the option – not offered in 'regular' Taguchi methods – not to obtain the complete set of response values: for example, if only four effects are needed, one could opt to collect as few as five observations out of an eight-run experiment. (Of course six, seven, or eight can also be collected if desired.)

For ease of reference, a lean design can be represented by the notation $L_n 2^{k-p}$, where 2^{k-p} has the usual meaning and can be replaced by 2^k when a full factorial is used, and n represents the number of experimental runs intended to be actually carried out. Note that $n \leq 2^{k-p}$. Since E_0 always has to be provided for, n is at least equal to one more than the total number of main effects and interaction effects requested by the investigator.

19.4 POTENTIAL OF LEAN DESIGN

For an appreciation of the potential of lean design, the case of the eight-run experiment is fully explored in Table 19.1. For example, with four factors, a requirement for four main effects but with total disregard for interaction effects can be met with a design of type $L_5 2^{4-1}$. For a regular eight-run 2^{5-2} design for five factors, if no interactions are needed, an $L_6 2^{5-2}$ design can be used instead. Similarly, for a 2^{6-3} design for six factors, a design of type $L_7 2^{6-3}$ is available. Considerable savings in experimental

Table 19.1 Some lean design possibilities for a regular eight-run design.

No. of factors	Total no. of necessary main and interaction effects	Smallest lean design available	Reduction in no. of runs.	Reduction in experimental resources (%)
3	4	$L_5 2^3$	3	37.5
	5	$L_6 2^3$	2	25.0
	6	$L_7 2^3$	1	12.5
	7	$L_8 2^3$	0	–
4	4	$L_5 2^{4-1}$	3	37.5
	5	$L_6 2^{4-1}$	2	25.0
	6	$L_7 2^{4-1}$	1	12.5
	7	$L_8 2^{4-1}$	0	–
5	4	$L_5 2^{5-2}$	3	37.5
	5	$L_6 2^{5-2}$	2	25.0
	6	$L_7 2^{5-2}$	1	12.5
	7	$L_8 2^{5-2}$	0	–
6	4	$L_5 2^{6-3}$	3	37.5
	5	$L_6 2^{6-3}$	2	25.0
	6	$L_7 2^{6-3}$	1	12.5
	7	$L_8 2^{6-3}$	0	–
7	4	$L_5 2^{7-4}$	3	37.5
	5	$L_6 2^{7-4}$	2	25.0
	6	$L_7 2^{7-4}$	1	12.5
	7	$L_8 2^{7-4}$	0	–

effort (37.5%, 25.0% and 12.5%, respectively) can be realized. Note that all the eight-run lean designs, that is, $L_8 2^{k-p}$, are no different from regular 2^{k-p} designs, and do not lead to any experimental data reduction.

Generally, the savings in experimental effort can be deduced as follows. Given a regular design of 2^{k-p} runs, n' main and interaction effects may be asked for:

$$n' < n \leq 2^{k-p}. \tag{19.1}$$

The maximum reduction in number of observations in a lean design $L_n 2^{k-p}$ is then

$$n_r = 2^{k-p} - n' - 1. \tag{19.2}$$

An illustration of calculations in a lean design experiment is now in order. Suppose in a fractional factorial experiment for five factors, each response takes about a week to obtain, and the project manager considers 8 weeks too long for the investigation. A lean design of type $L_6 2^{5-2}$ can be used to cut data collection time by a fortnight, supported by the technical judgment that interaction effects in this experiment are negligible.

In this situation, a standard 2^{5-2} design matrix can be first constructed based on part of the 2^{7-4} matrix shown in Table 19.2. Suppose columns x_6 and x_7 are not used in the experimental design. One can opt to omit two of the eight experimental runs i, $i = 1, 2, \ldots, 8$, as only six degrees of freedom are needed to calculate E_0 and five main

Table 19.2 Design matrix used in illustrations of lean design.

Run i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_i	Time (min)
1	−	−	−	+	+	+	−	y_1	68.4
2	+	−	−	−	−	+	+	y_2	77.7
3	−	+	−	−	+	−	+	y_3	66.4
4	+	+	−	+	−	−	−	y_4	81.0
5	−	−	+	+	−	−	+	y_5	78.6
6	+	−	+	−	+	−	−	y_6	41.2
7	−	+	+	−	−	+	−	y_7	68.7
8	+	+	+	+	+	+	+	y_8	38.7

effects. Let y_7 and y_8 be the two response values omitted as a result. Since columns x_6 and x_7 are idle, they correspond to the two zero effects E_6 and E_7. E_6 and E_7 in turn can be expressed in terms of the eight response values, as is done in a normal analysis of experimental data:

$$E_6 = \tfrac{1}{4}(y_1 + y_2 - y_3 - y_4 - y_5 - y_6 + \hat{y}_7 + \hat{y}_8) = 0, \tag{19.3}$$

$$E_7 = \tfrac{1}{4}(-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - \hat{y}_7 + \hat{y}_8) = 0, \tag{19.4}$$

where $y_1, y_2, \ldots, y_6$ are observed values, and $\hat{y}_7$, $\hat{y}_8$ are dummy values standing in for the two absent responses. Equations (19.3) and (19.4) can be solved simultaneously to yield

$$\hat{y}_7 = -y_2 + y_3 + y_5, \tag{19.5}$$

$$\hat{y}_8 = -y_2 + y_4 + y_6. \tag{19.6}$$

These two expressions, together with y_i, $i = 1, 2, \ldots, 6$, can now be used afresh to calculate E_j, $j = 0, 1, \ldots, 5$, in the usual manner to complete the required analysis.

It is important to note that when there is to be more than one absent response value, the choice of the omitted runs should be such that independent estimates of these values can be obtained. For example, if runs 2 and 8 are omitted, then

$$E_6 = \tfrac{1}{4}(y_1 + \hat{y}_2 - y_3 - y_4 - y_5 - y_6 + y_7 + \hat{y}_8) = 0, \tag{19.7}$$

$$E_7 = \tfrac{1}{4}(-y_1 + \hat{y}_2 + y_3 - y_4 + y_5 - y_6 - y_7 + \hat{y}_8) = 0, \tag{19.8}$$

which will not lead to unique solutions for $\hat{y}_2$ and $\hat{y}_8$. This precaution can be taken by simply inspecting the full 2^{7-4} design matrix shown in Table 19.2.

Lean design can also be used in a situation where the intention is to eventually complete a regular design, say, of eight runs, but it is possible to exercise some technical judgment to assume the relative unimportance of certain effects – so as to ignore them in the beginning in exchange for the opportunity to skip some response values and come to some quick preliminary conclusions. Thereafter, additional experimental runs can be carried out to make good the absent response values, for a final regular

eight-run analysis. It may be noted that to do this, some caution is to be exercised to avoid any possible block effect associated with the sequential data collection.

19.5 ILLUSTRATIVE EXAMPLE

For an illustration of such an approach, consider the frequently cited filtration plant example on pp. 424–429 of Box *et al*.[6] In this example, various factors were suggested for an investigation of the cycle time at a filtration plant: source of water supply (x_1), original of raw material (x_2), level of temperature (x_3), presence of recycling (x_4), rate of addition of caustic soda (x_5), type of filter cloth (x_6), and length of hold-up time (x_7). The experimental design used was 2^{7-4}, as shown in Table 19.2, in which the response values are also shown. The results of the analysis are as follows, where l_j, $j = 1, 2, \ldots, 7$, are, respectively, values of E_j, $j = 1, 2, \ldots, 7$, each confounded with interaction effects:

$$l_1 = E_1 + E_{24} + E_{35} + E_{67} = -10.875,$$
$$l_2 = E_2 + E_{14} + E_{36} + E_{57} = -2.775,$$
$$l_3 = E_3 + E_{15} + E_{26} + E_{47} = -16.575,$$
$$l_4 = E_4 + E_{12} + E_{37} + E_{56} = 3.175,$$
$$l_5 = E_5 + E_{13} + E_{27} + E_{46} = -22.825,$$
$$l_6 = E_6 + E_{17} + E_{23} + E_{45} = -3.425,$$
$$l_7 = E_7 + E_{16} + E_{25} + E_{34} = 0.525.$$

Based on a Pareto analysis, it was concluded that the important effects were l_5, l_1 and l_3, and attention was focused on x_5, x_1 and x_3 as prime candidates for factors crucial to the filtration time variation. Another 2^{7-4} experiment based on the fold-over technique finally identified x_5 and x_1 as the factors that really mattered.

During the initial phase of the investigation, it was already suspected by the person responsible that out of the seven factors, perhaps only one or two, at most three, factors would be found important. That being the case, if there had to be a constraint on the investigation such as time, then lean design would be extremely helpful. Suppose there was only time for six experimental runs before the first results of the investigation were due; the investigator could use his best judgment to name two factors least likely to be important to the process. These factors would still be included in the experimental work, but now only six instead of eight observations would need to be taken in the first instance: in this way a 25% reduction in experimentation time was possible prior to the first interim report on the investigation.

Consider an $L_6\ 2^{7-4}$ lean design comprising the first six runs listed in Table 19.2; that is to say two response values, y_7 and y_8, are absent. Assume that x_6 and x_7 are considered by the investigator to be least important in the filtration process, so that l_6 and l_7 can be set to zero. Then, using the reasoning explained earlier, y_7 and y_8 can be estimated by equations (19.5) and (19.6) to be 76.6 and 44.5, respectively. With these values, the l_j values, $j = 1, 2, \ldots, 5$, can be computed.

Table 19.3 Results of lean designs in incremental experimentation.

Estimated effect	$L_6 2^{7-4}$ y_7, y_8 absent	$L_7 2^{7-4}$ y_8 absent	2^{7-4} Regular data set
l_1	−11.40	−11.40	−10.875
l_2	0.65	−3.30	−2.775
l_3	−13.15	−17.10	−16.575
l_4	2.65	2.65	3.175
l_5	−23.35	−23.35	−22.825
l_6	0	−3.95	−3.425
l_7	0	0	0.525

Another possible lean design is one with seven experimental runs, or $L_7 2^{7-4}$. In this case only one factor needs to be assumed unimportant. If x_7 is such a factor and y_8 is unavailable, then one can set

$$l_7 = \tfrac{1}{4}(-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + \hat{y}_8) = 0. \tag{19.9}$$

Hence

$$\hat{y}_8 = y_1 - y_2 - y_3 + y_4 - y_5 + y_6 + y_7 = 36.6, \tag{19.10}$$

with which all the other l_i values can be computed.

The results of analysis of the above two lean designs are exhibited in Table 19.3, along with the results obtained from the full 2^{7-4} experimental data from the filtration study. It is evident from the table that, using an $L_6 2^{7-4}$ design, the investigator would arrive at exactly the same four tentative conclusions associated with the data from the regular 2^{7-4} design (p. 426 of Box *et al.*[6]). With one more experimental run – that is, with an $L_7 2^{7-4}$ design – there would still be no change in the interpretation. This shows, in a striking manner, that with a considerably smaller amount of experimental effort, lean design works extremely well when the investigator is able to make a good prior judgment concerning the relative importance – or rather, unimportance – of the factors under study.

Even if the judgment of the investigator is somewhat off the mark, especially in the presence of unsuspected interactions, the results of a lean design can still be gradually improved upon by the addition of one further experimental run at a time. Thus the filtration plant study can start with a lean design of $L_6 2^{7-4}$ to obtain some provisional findings, followed by the generation of response y_7 to produce the results corresponding to an $L_7 2^{7-4}$ design, and finally y_8 to arrive at what a regular 2^{7-4} design can reveal about the plant behavior. Such an *incremental experimentation* or one-additional-run-at-a-time scheme, has the same merit as that advocated for sequential fold-over designs, namely earlier availability of first results, and improvement of such results as new data becomes available.

19.6 POSSIBLE APPLICATIONS

Lean design is useful in circumstances such as those described below.

1. It is imperative to have only the minimum number of experimental runs in view of, for example, expensive prototypes, high costs of destructive tests, or costly disruptions to regular production.

2. It is foreseen that constraints in certain experimental resources, for example materials, equipment or manpower availability, are such that there will be long delays if one has to await a complete set of data associated with a regular experimental design.
3. Some quick answers concerning the subject of study are urgently needed, hence one cannot afford the time for a full experimental data set.
4. Sometimes it may be deemed important to have information on experimental errors; without increasing the total number of observations, replication can be made in certain experimental runs in return for absence of response values in others, a trade-off realizable via lean design.
5. It is envisaged that a certain condition specified in an experimental design matrix, such as 'high' (+1) settings for all the factors, may adversely affect a process, create hazardous conditions, or cause undue inconvenience on the production floor, and should be avoided: the response corresponding to such a condition may have to be left out on purpose.

   The above would mean that incomplete experimentation based on a lean design is intentional. In situations where an experiment is originally meant to be run in full, lean design may be invoked to substitute the original design, as illustrated by the following.
6. The infeasibility of an experimental condition – for a reason similar to that in case 5 – is discovered only during the course of data collection, so arrangements for a particular response have to be abandoned.
7 The experiment has to be aborted owing to unforeseen circumstances, such as equipment breakdown, exhaustion of raw materials, or unavailability of personnel, and the best has to be made of the data already collected.
8. Results of the study are suddenly called for before the completion of experiment, to be used to help deliberations in an urgent meeting, for instance.
9. After an experiment is completed, it is found that certain data are associated with faulty parts and materials, or have been wrongly measured or otherwise corrupted and hence cannot be used, and it is not feasible to re-enact the spoilt runs.

Regardless of the circumstances in which lean design is used, if there is subsequently an opportunity to make up for the absent response values, the additional data can always be combined with those already available for a fresh analysis of the experiment. This in fact also provides an opportunity to check on the validity of initial assumptions concerning the non-existence of certain interaction effects. In this way an incremental improvement of the experimental results is possible with the availability of each additional late response value. When this is intentionally done, for example using $L_5 2^{4-1}$ as a start, then including one additional response value at a time for a re-analysis of the experiment – moving through $L_6 2^{4-1}$, $L_7 2^{4-1}$ and $L_8 2^{4-1}$ designs until eight values are finally available – is in fact execution of an incremental experimentation scheme.

The use of lean design for incremental experimentation, depicted by the flow chart in Figure 19.1, would prove useful in situations where every bit of early information is crucial, for example in research and development for new advanced materials, or urgent troubleshooting of high-cost processes. Not only does the approach incur no more resources than called for in a regular experimental design, its *self-correction*

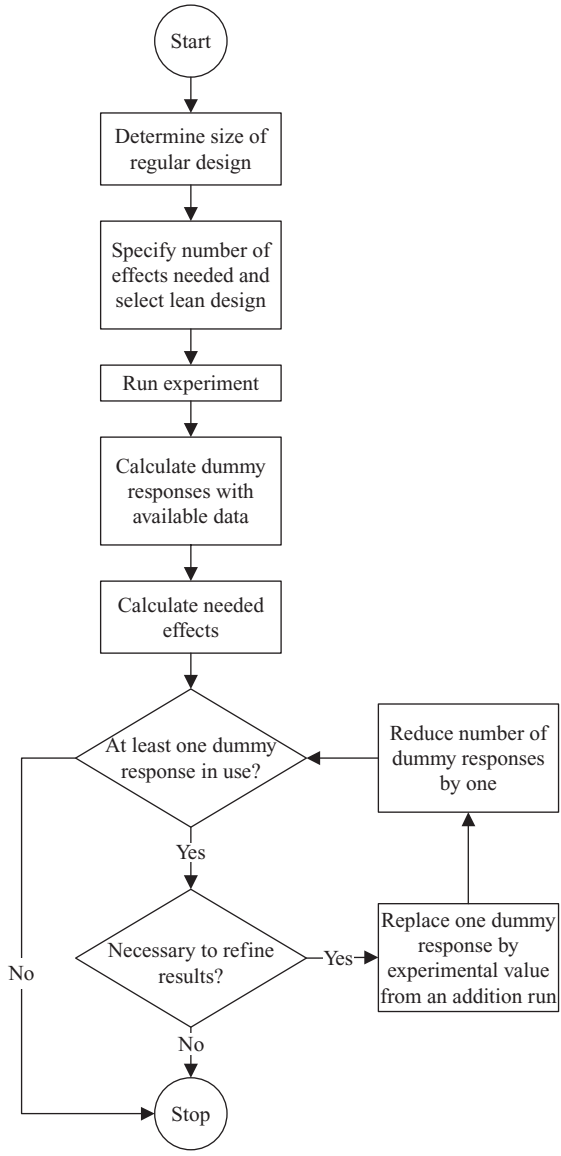

Start
Determine size of regular design
Specify number of effects needed and select lean design
Run experiment
Calculate dummy responses with available data
Calculate needed effects
At least one dummy response in use?
Reduce number of dummy responses by one
Yes
Necessary to refine results?
Yes
Replace one dummy response by experimental value from an addition run
No
No
Stop

Figure 19.1 Lean design for incremental experimentation.

nature inherent in the results generated at each stage is also consonant with the spirit of 'continuous improvement' in an investigation.

19.7 CONCLUDING REMARKS

Experimental design is normally understood to be used in such a way that, regardless of the type and size of experiment, no data analysis could commence before the full set of response values is available. In other words, the investigator must wait for the

very last observation from his experiment before he is in a position to announce any findings at all from his work. Situations abound in industry where one may not have the luxury of awaiting the full data set, owing to foreseen or unforeseen circumstances similar to those highlighted in the previous section. In many cases, rapid diagnosis of a troubled process is the best way to retain customer confidence, and in other settings, early availability of indicative results is vital to competitiveness in research and development: the time taken to generate full experimental data sets may simply be unacceptable. Lean design can provide a way out, as long as the user is fully aware of the possible pitfalls associated with the no-interaction assumption typically seen in a Taguchi-type design – and, better still, aided by prior knowledge concerning the relative strengths of the factors under investigation. With this proviso, there are several advantages that would make lean design attractive to both management and technical personnel:

1. A lean design can be specified at the outset to suit given time and resource budgets.
2. A lean design takes maximum advantage of any prior technical knowledge about the significance or otherwise of main and interaction effects, saving investigation costs through reduction of experimental runs.
3. The use of lean design and incremental experimentation fits the Taguchi school of thought regarding the handling of interaction effects, yet makes use of mainstream statisticians' favored approach of acquiring and updating information via sequential data gathering.
4. Although quick results from lean experiments could be less reliable than results associated with regular orthogonal designs, they can be subsequently improved upon – in fact, self-correct – by incremental experimentation, to a point that is identical to the results of regular designs. This, indeed, would be the clinching argument for using lean design as the first stage in an experimental project.

REFERENCES

1. Kacker, R.N. (1985) Off-line quality control, parameter design, and the Taguchi method. *Journal of Quality Technology*, 17, 176–188.
2. Goh, T.N. (1993) Taguchi methods: some technical, cultural and pedagogical perspectives. *Quality and Reliability Engineering International*, 9, 185–202. See also Chapter 18, this volume.
3. Box, G.E.P., Bisgaard, S., and Fung, C.A. (1988) An explanation and critique of Taguchi's contributions to quality engineering. *Quality and Reliability Engineering International*, 4, 123–131.
4. Pignatiello, J.J. and Ramberg, J.S. (1991) Top ten triumphs and tragedies of Genichi Taguchi. *Quality Engineering*, 4, 211–225.
5. Goh, T.N. (1994) Taguchi methods in practice: an analysis of Goh's paradox. *Quality and Reliability Engineering International*, 10, 417–421.
6. Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978) *Statistics for Experimenters*. New York: John Wiley & Sons, Inc.

20

A Unified Approach for Dual Response Surface Optimization

L. C. Tang and K. Xu

The optimization of dual response systems in order to achieve better quality has been an important problem in the robust design of industrial products and processes. In this chapter, we propose a goal programming approach to optimize a dual response system. The formulation is general enough to include some of the existing methods as special cases. For purposes of illustration and comparison, the proposed approach is applied to two examples. We show that by tuning the weights assigned to the respective targets for the mean and standard deviation, past results can easily be replicated using Excel Solver, the use of which further enhances the practical appeal of our formulation.

20.1 INTRODUCTION

The dual response surface approach, first considered by Myers and Carter[1] and revitalized by Vining and Myers,[2] is important in determining the optimal operating conditions for an industrial process, so that the target value for the mean of a quality characteristic is met and its variability is also minimized. Using the approach, some of the goals of Taguchi's philosophy can be realized without resorting to combining information about both mean and variability into a signal-to-noise ratio.

Basically, in the dual response surface approach, two empirical models, one for the mean and the other for the standard deviation of the response, are first established, and then the two fitted response surface models are optimized simultaneously in a region of interest ($\mathbf{x} \in \Omega$). Typical second-order models for the mean and

This chapter is based on the article by L. C. Tang and K. Xu, 'A unified response for dual response surface optimization', *Journal of Quality Technology*, 34(4), 2002, pp. 437–447, and is reproduced by the permission of the publisher, the American Society for Quality

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

variance are:

$$\hat{y}_\mu = a_0 + \sum_{i=1}^{k} a_i x_i + \sum_{i=1}^{k} a_{ii} x_i^2 + \sum \sum_{i<j}^{k} a_{ij} x_i x_j,$$

$$\hat{y}_\sigma = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum \sum_{i<j}^{k} b_{ij} x_i x_j$$

where $\hat{y}_\mu$ is the estimated value of the mean and $\hat{y}_\sigma$ is the estimated value of the standard deviation.

Various optimization schemes have been proposed in past 15 years. Here, we propose a unified formulation, which includes some existing techniques as special cases. The approach will thus facilitate comparison and interpretation of results obtained via different techniques. Through the use of two examples from published work, we shall demonstrate the versatility of the proposed scheme by replicating results given by other researchers. The results are obtained by adjusting the weight parameters in the constraints so that the trade-off between meeting the targets for the dual responses can be explicitly incorporated. All computations are carried out in Excel Solver which can easily be implemented and be used by most practitioners.

This chapter is organized as follows. A review of existing techniques is presented in Section 20.2. This is followed by a description of the proposed optimization scheme. In Section 20.4, the method is tested and illustrated using two examples, one of which has been extensively investigated by various researchers. A detailed comparison and a sensitivity analysis are also presented. Finally, a brief conclusion is given.

20.2 REVIEW OF EXISTING TECHNIQUES FOR DUAL RESPONSE SURFACE OPTIMIZATION

Various optimization schemes have been proposed in the past 15 years. The scheme proposed by Vining and Myers[2] (referred to as VM) is to optimize a primary response subject to an appropriate secondary response constraint using the Lagrangian multiplier approach:

$$\min_X \text{ (or max)} \quad y_{\text{primary}}$$

$$\text{subject to} \quad y_{\text{secondary}} = \varepsilon,$$

where ε is a specific target value.

Del Castillo and Montgomery[3] (referred to as DM) solved the problem using the generalized reduced gradient algorithm, which is available in some software packages such as Excel Solver. It has been demonstrated that the approach can handle the cases of target-is-best, larger-is-better, and smaller-is-better.

Lin and Tu[4] (referred to as LT) noted that the optimization scheme based on Lagrangian multipliers may be unrealistic, since it forces the estimated mean to a specific value. They thus propose minimizing the mean squared error (MSE) instead:

$$\min_{X} \quad MSE = (\hat{y}_\mu - T)^2 + \hat{y}_\sigma^2.$$

Copeland and Nelson[5] (referred to as CN) pointed out that minimizing the MSE places no restriction on how far the resulting value of $\hat{y}_\mu$ might be from the target value T and suggested the use of direct minimization of an objective function. For target-is-best situation, the proposed scheme is to minimize the function $\hat{y}_\sigma + \varepsilon$, where

$$\varepsilon = \begin{cases} (\hat{y}_\mu - T)^2, & \text{if} \quad (y_\mu - T)^2 > \Delta^2, \\ 0, & \text{if} \quad (y_\mu - T)^2 \leq \Delta^2. \end{cases}$$

A quality lost function model proposed by Ames *et al.*[6] for multiple response surfaces can also be applied to the dual response case. The scheme is:

$$\min \; QLP = \sum_{i=1}^{m} \omega_i \, (\hat{y}_i - T_i)^2,$$

where $\hat{y}_i$ is the estimated response for quality criterion i, with respective target value, T_i, and W_i is the corresponding weight for each response.

Recently, Kim and Lin[7] (referred to as KL) proposed a fuzzy approach for dual response surface optimization based on the desirability function formulation,[8] a popular multiresponse optimization technique. Vining[9] extended the work of Khuri and Conlon,[10] Pignatiello,[11] and Ames *et al.*[6] and proposed an optimization scheme using a squared error loss function that explicitly incorporates both the variance–covariance structure of both of the responses.

20.2.1 Optimization scheme

For a dual response system in which targets can be specified, it is natural to try to achieve the ideal targets as closely as possible. While there are several ways to achieve such objective, we propose a formulation using the idea of goal programming,[12,13] in which both the objective function and constraints are of quadratic form as follows:

$$\begin{aligned} &\min \quad \delta_\mu^2 + \delta_\sigma^2 \\ &\text{subject to} \quad y_\mu(\mathbf{x}) - \omega_\mu \delta_\mu = T_\mu^*, \qquad y_\sigma(\mathbf{x}) - \omega_\sigma \delta_\sigma = T_\sigma^*, \\ &\text{and either} \quad \mathbf{x}\mathbf{x}^{\mathrm{T}} \leq r^2 \quad \text{or} \quad \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u. \end{aligned} \tag{20.1}$$

The function to be minimized $\delta_\mu^2 + \delta_\sigma^2$, is the objective function where $\delta_\mu, \delta_\sigma$ are unrestricted scalar variables (i.e. they can be positive or negative). The subsequent two equations are the constraints where $\omega_\mu, \omega_\sigma$ ($\geq$0) are user-defined weights; $\mathbf{T}^* = \{T_\mu^*, T_\sigma^*\}$ are the ideal response mean and standard deviation associated with a set of response functions $\hat{\mathbf{Y}}(\mathbf{x}) = \{\hat{y}_\mu(\mathbf{x}), \hat{y}_\sigma(\mathbf{x})\}$. which are typically of quadratic form; and $\mathbf{x} = [x_1, x_2, \ldots, x_a]^{\mathrm{T}}$ is set of control factors. The typical solution space of control factors that is of interest can be either $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq r^2$ for spherical region, where r is the

radius of the zone of interest, or $\mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u$ for rectangular region. The products $\omega_i \delta_i$ $(i = \mu, \sigma)$ introduce an element of slackness into the problem, which would otherwise require that the targets T_i $(i = \mu, \sigma)$ be rigidly met; which will happen if ω_μ and ω_σ are both set to zero. The setting of $\omega_\mu, \omega_\sigma$ will thus enable the user to control the relative degree of under- or over-achievement of the targets for the mean and standard deviation, respectively. The relative magnitudes of the ω_i also represents the user's perception of the relative ease with which the optimal solution of y_i $(\mathbf{x})$ will achieve the targets T_i^*. By using different parameter values, alternative Pareto solutions with varying degrees of under-attainment can be found. A Pareto solution[14] (also called a noninferior or efficient solution) is one in which an improvement in the meeting of one target requires a degradation in the meeting of another.

The above formulation can be run in Excel Solver using a simple template shown in Figure 20.1. In general, it is recommended that the optimization program be run with different starting points. Although the scheme may not guarantee a global optimum, it is a simple and practical technique that gives good results that are close to the global optimal solution in many instances. The issues involved in using Solver are discussed by Del Castillo and Montgomery.[3] An algorithm (DRSALG) and its ANSI FORTRAN implementation that guarantees global optimality, in a spherical experimentation region, are presented by Del Castillo *et al.*[15,16] Recently, Fan[17] presented an algorithm, DR2, that can guarantee a global optimal solution for nondegenerate problems and

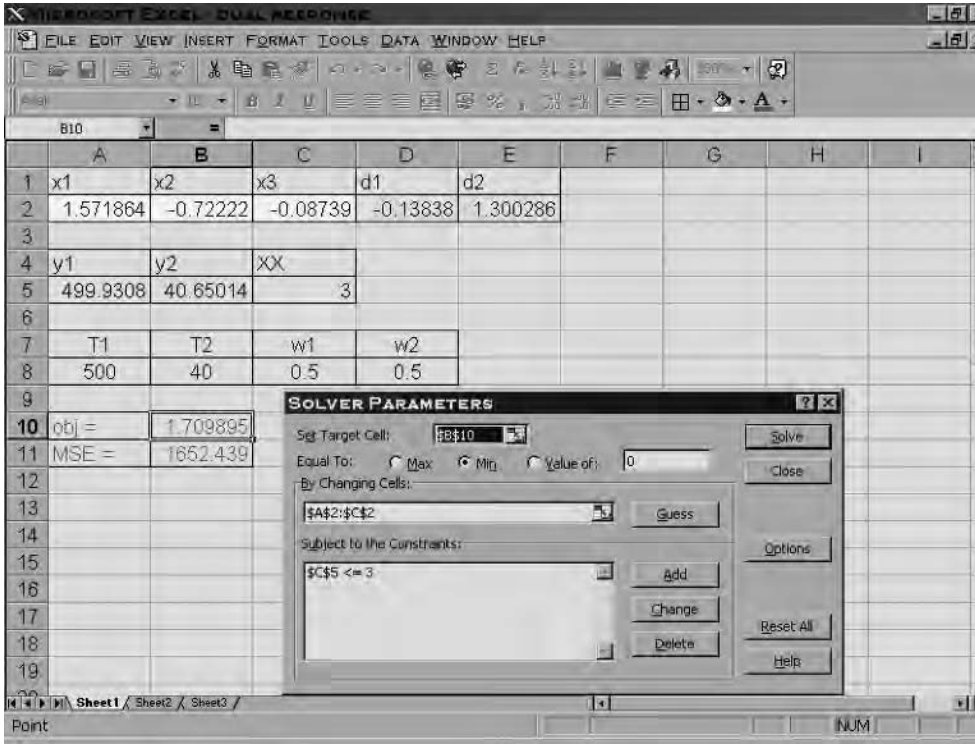


Figure 20.1 Screen shot for implementing the proposed formulation in Excel Solver.

returns a near-global one for degenerate problems within a radial region of experimentation. A comprehensive discussion and comparison of various algorithms are also given in the papers cited.

20.2.2 Comparison of proposed scheme with existing techniques

The proposed method provides a general framework that unifies some of the existing techniques. This can be seen as follows.

Consider LT's formulation [4] in which the MSE is minimized:

$$\min_X \quad MSE = (\hat{y}_\mu - T)^2 + \hat{y}_\sigma^2. \tag{20.2}$$

From equation (20.1), if we set $\omega_\mu = \omega_\sigma = 1$ and $T_\sigma^* = 0$, we have

$$\begin{aligned} &\min \quad \delta_\mu^2 + \delta_\sigma^2 \\ &\text{subject to } \hat{y}_\mu(\mathbf{x}) - \delta_\mu = T_\mu^*, \qquad \hat{y}_\sigma(\mathbf{x}) - \delta_\sigma = 0. \end{aligned} \tag{20.3}$$

It follows that $\delta_\mu = \hat{y}_\mu(\mathbf{x}) - T_\mu^*$ and $\delta_\sigma = y_\sigma(\mathbf{x})$. Thus equation (20.3) is equivalent to (20.2), since

$$\delta_\mu^2 + \delta_\sigma^2 = \left(\hat{y}_\mu(\mathbf{x}) - T_\mu^*\right)^2 + y_\sigma(\mathbf{x})^2.$$

Next, consider VM and DM's target-is-best formulation,[2,3] which can be formulated as

$$\begin{aligned} &\min \qquad \hat{y}_\sigma \\ &\text{subject to} \quad \hat{y}_\mu = T_\mu^*. \end{aligned} \tag{20.4}$$

From (20.1), if we set $\omega_\mu = 0$, $\omega_\sigma = 1$, and $T_\sigma^* = 0$, then δ_μ has no bearing on (20.1) and $\delta_\sigma = \hat{y}_\sigma(\mathbf{x})$. It follows that (20.1) is equivalent to

$$\begin{aligned} &\min \qquad \hat{y}_\sigma^2 + \text{constant} \\ &\text{subject to} \quad \hat{y}_\mu(\mathbf{x}) = T_\mu^*, \end{aligned}$$

which is the same as (20.4).

It is clear that by using suitable settings for ω_μ, ω_σ, and T_i^*, other existing formulations can be replicated. These are summarized in Table 20.1, where $T_\mu^{\max}$ and $T_\mu^{\min}$ are respectively the maximum and minimum possible values for the mean response so that we have $\min\left(\hat{y}_\mu - T_\mu^{\max}\right)^2 \equiv \max(\hat{y}_\mu)$ and $\min\left(\hat{y}_\mu - T_\mu^{\min}\right)^2 \equiv \min(\hat{y}_\mu)$. In all the above schemes, the solution space is restricted to $\mathbf{x}\mathbf{x}^{\mathrm{T}} \le r^2$ or $\mathbf{x}_l \le \mathbf{x} \le \mathbf{x}_u$.

Although an equivalent formulation for the CN approach[5] cannot be readily obtained from the proposed scheme, we shall show in the numerical example that identical result can be obtained by tuning the weights. It is worth noting that our scheme not only provides the slackness for the mean target as in LT and CN, but also includes the slackness from the variance response target, which others do not consider. We discuss later how the trade-off between meeting the mean and variance targets can be expressed explicitly by tuning the weights. Such flexibility, which is not available in previous approaches, enhances the practical appeal of the current formulation.

Table 20.1 Comparison of proposed scheme with existing techniques.

Case	Existing optimization scheme	Proposed scheme
1.	LT: Lin and Tu[4] $\min_{X} \left(\hat{y}_\mu - T_\mu^*\right)^2 + \hat{y}_\sigma^2$	$\omega_\mu = 1,\ \omega_\sigma = 1\,,\ T_\sigma^* = 0$ $\Rightarrow \delta_\mu = y_\mu(\boldsymbol{x}) - T_\mu^*, \delta_\sigma = \hat{y}_\sigma(\boldsymbol{x})$ $\Rightarrow \delta_\mu^2 + \delta_\sigma^2 = \left(\hat{y}_u(\boldsymbol{x}) - T_\mu^*\right)^2 + \hat{y}_\sigma(\boldsymbol{x})^2$
2.	VM and DM: Target-is-best[2,3] $\min\ \hat{y}_\sigma$ subject to $\hat{y}_\mu = T_\mu^*$	$\omega_\mu = 0,\ \omega_\sigma = 1\,,\ T_\sigma^* = 0$ $\Rightarrow \delta_\sigma = \hat{y}_\sigma,$ $\min\ \hat{y}_\sigma^2$ subject to $\hat{y}_\mu = T_\mu^*$
3.	VM and DM: Larger-is-better[2,3] $\max\ \hat{y}_\mu$ subject to $\hat{y}_\sigma = T_\sigma^*$	$\omega_\mu = 1, \omega_\sigma = 0, T_\mu^* = T_\mu^{\max}$ $\Rightarrow \delta_\mu = \hat{y}_\mu - T_\mu^{\max},$ $\min\ \left(\hat{y}_\mu - T_\mu^{\max}\right)^2$ subject to $y_\sigma = T_\sigma^*$
4.	VM and DM: Smaller-is-better[2,3] $\min\ \hat{y}_\mu$ subject to $\hat{y}_\sigma = T_\sigma^*$	$\omega_\mu = 1, \omega_\sigma = 0, T_\mu^* = T_\mu^{\min}$ $\Rightarrow \delta_\mu = y_\mu - T_\mu^{\min}$ $\min\ \left(y_\mu - T_\mu^{\min}\right)^2$ subject to $y_\sigma = T_\sigma^*$
5.	Quality lost function model[6] $\min\ QLP = \sum_{i=1}^{m} \omega_i \left(\hat{y}_i - T_i\right)^2$	The constraints can be written as $\delta_\mu = \left(\frac{\hat{y}_\mu - T_\mu^*}{\omega_\mu}\right)$ and $\delta_\sigma = \left(\frac{\hat{y}_\sigma - T_\sigma^*}{\omega_\sigma}\right)$ for $\omega_\mu, \omega_\sigma \neq 0$ $\min\ \left(\frac{\hat{y}_\mu - T_\mu^*}{\omega_\mu}\right)^2 + \left(\frac{\hat{y}_\sigma - T_\sigma^*}{\omega_\sigma}\right)^2$

It can also be seen from Table 20.1 that the three common goals, the target-is-best, the larger-is-better and the smaller-is-better, are represented by cases 2 through 4, respectively.

20.2.3 Setting ω_μ and ω_σ

Besides establishing the dual response surface for the problem under consideration, a crucial step in implementing the proposed scheme is the setting of the values of ω_μ and ω_σ. Readers are referred elsewhere for the construction of the dual response surfaces.[18–21] Here, we focus on the choosing the weights ω_μ and ω_σ.

The weights, in some sense, reflect a practitioner's preference for the closeness of the mean and standard deviation of the response to their respective target values. The interpretations of several possible schemes are as follows (for the justifications, see Table 20.1).

- Set $\omega_\mu = \omega_\sigma = 1$. This scheme will lead to minimizing the MSE since the sum of the square of the distance measures of the mean and variance responses from their

respective targets, i.e. the true mean and zero, are minimized. It is in fact the same objective as in Lin and Tu.[4]

- Set $\omega_\mu = 0, \omega_\sigma = 1, T_\sigma^* = 0$. This scheme is reduced to the target-is-best formulation.[2,3]
- Setting the $\omega_\mu = 1, \omega_\sigma = 0, T_\mu^* = T_\mu^{\max}$, This scheme is reduced to the larger-is-better formulation.[2,3]
- Set $\omega_\mu = 1, \omega_\sigma = 0, T_\mu^* = T_\mu^{\min}$. This scheme is reduced to the smaller-is-better formulation.[2,3]

In order to assess the effect of the weights on the final solution, a sensitivity analysis should be performed. This is best done through experimenting with different convex combinations of the weights, that is, $\omega_\mu + \omega_\sigma = 1, \omega_\mu > 0, \omega_\sigma > 0$. In this way, ω_μ and ω_σ can be varied parametrically to generate a set of noninferior solutions for this multiple objective optimization problem (see the weighting method in Cohon[22] and Ehrgott[23]).

Note that for weights between 0 and 1, our formulation is similar to that of the quality loss model proposed by Ames *et al.* [6] Thus, the choice of weights can be tuned to represent the preference for quality loss.

In addition, there are two interesting settings of the weights that could also represent the preference of a decision maker.

- Set the weights equal to their corresponding targets. This results in

$$\delta_\mu = \frac{\hat{y}_\mu(\mathbf{x}) - T_\mu^*}{T_\mu^*} \quad \text{and} \quad \delta_\sigma = \frac{\hat{y}_\sigma(\mathbf{x}) - T_\sigma^*}{T_\sigma^*}.$$

  Here, $\delta_\mu, \delta_\sigma$ actually represent the percentage of under- or over-attainment from the targets for the dual responses. Thus minimizing $\delta_\mu^2 + \delta_\sigma^2$ is identical to minimizing the percentage of deviation from respective targets with the same preference.
- Set the weights equal to the reciprocal of the targets. From equation (20.1), the constraints become

$$\delta_\mu = \left(\hat{y}_\mu(\mathbf{x}) - T_\mu^*\right) T_\mu^* \quad \text{and} \quad \delta_\sigma = \left(\hat{y}_\sigma(\mathbf{x}) - T_\sigma^*\right) T_\sigma^*.$$

  Thus,

$$\min\ \left(\delta_\mu^2 + \delta_\sigma^2\right) = \min\left[\left(\hat{y}_\mu(\mathbf{x}) - T_\mu^*\right)^2 T_\mu^{*2} + \left(\hat{y}_\sigma(\mathbf{x}) - T_\sigma^*\right)^2 T_\sigma^{*2}\right].$$

  The objective function becomes that of minimizing the weighted sum of squares of deviations from the targets in which the weights are given by the square of the respective target value. This will result in a smaller deviation for the larger target value due to higher weight. It thus implies the preference for smaller deviation from the larger target value.

20.3 EXAMPLE 1

In this section, we compare various solutions from our formulation with other results using a classical example of a printing process. This example is taken from Box and Draper,[18] and it is also used elsewhere.[2–5,7] The purpose of the experiment is to analyze the effect of the speed, pressure, and distance variables on the ability that a printing machine has for applying colored ink to package labels. The proposed fitted response surface for the mean and standard deviation of the characteristic of interest are as follows:[2]

$$\hat{y}_\mu = 327.6 + 177.0x_1 + 109.4x_2 + 131.5x_3 + 32.0x_1^2 - 22.4x_2^2 - 29.1x_3^2 + 66.0x_1x_2 + 75.5x_1x_3 + 43.6x_2x_3,$$

$$y_\sigma = 34.9 + 11.5x_1 + 15.3x_2 + 29.2x_3 + 4.2x_1^2 - 1.3x_2^2 + 16.8x_3^2 + 7.7x_1x_2 + 5.1x_1x_3 + 14.1x_2x_3.$$

20.3.1 Target-is-best

We consider two cases with the target for the mean at 500, with different targets for the standard deviation at 40 and 44.72.

- There is a circular region of interest, $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$. In Table 20.2, the ideal target employed in this case is (500, 40) and the effects of several weights combinations are given. For comparison, the results from DM, LT, and CN are given in Table 20.3, together with the weights that would give the same result under our scheme.
- There is a cubical region of interest, $-\mathbf{1} \leq \mathbf{x} \leq \mathbf{1}$. The solution of the proposed approach is shown in Table 20.4 for this region of interest. The results of various existing techniques for the print process example within this region of interest are shown in Table 20.5.

It can be seen from Table 20.4 that the results of the proposed method with weights (1/500, 1/44.72) are close to those obtained by DM. Note that, in Table 20.5, for the LT and KL approaches where the MSE is minimized, the weights are almost equal. This

Table 20.2 Results of the proposed approach for target $\mathbf{T}^* = (500, 40)$ ($\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$).

$\mathbf{T}^*$	ω $(\omega_\mu, \omega_\sigma)$	$\mathbf{Y}(\mathbf{X})^{\mathrm{T}}$ (y_μ, y_σ)	$\mathbf{X}^{\mathrm{T}}$ (x_1, x_2, x_3)	MSE	Delta $(\delta_\mu, \delta_\sigma)$
(500, 40)	(1/500, 1/40)	(499.9996, 40.6575)	(1.5718, −0.7224, −0.0867)	1653.03	(−0.2239, 26.2987)
(500, 40)	(1, 1)	(499.9308, 40.6502)	(1.5717, −0.7226, −0.0868)	1652.44	(−0.0692, 0.6501)
(500, 40)	(500, 40)	(496.0527, 40.2379)	(1.5664, −0.7342, −0.0863)	1634.67	(−0.0079, 0.0059)

Table 20.3 Results of existing methods ($\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$).

Methods	$Y(\mathbf{x})^{\mathrm{T}}$ $(\gamma_\mu, \gamma_\sigma)$	$\mathbf{X}^{\mathrm{T}}$ (x_1, x_2, x_3)	MSE	Proposed Scheme $(\omega_\mu, \omega_\sigma)$
CN	(495.88, 40.22)	(1.5683, −0.7290, −0.0946)	1634.68	(0.49, 0.51)
VM and DM	(500, 40.65)	(1.5719, −0.7220, −0.0874)	1652.46	(0, 1)
LT	(495.68, 40.20)	(15651, −0.7373, −0.0883)	1634.57	(0.5, 0.5)

supports our interpretation of weight selection. More detailed results and discussions are presented next by exploring different convex combinations of the weights.

20.3.2 Sensitivity analysis

Note that in the above analysis, the targets and weights are based on practical considerations and the users' preferences discussed earlier. In order to have a clearer picture of the influence of the weights on the solution obtained, a sensitivity analysis is presented. Here we select the target as (500, 0) and use convex combinations of the weights, $\omega_\mu^j + \omega_\sigma^j = 1$, $j = 1, \ldots, m$. Using an iterative scheme in which $\omega_\mu^{j+1} = \omega_\mu^j + \Delta\omega$, $\omega_\sigma^{j+1} = \omega_\sigma^j - \Delta\omega$, with $\Delta\omega = 0.01$, a set of noninferior solutions is generated. Partial results are given in Tables 20.6 and 20.7 for the two regions of interest.

It is interesting to note that all the results obtained by other methods such as CN, DM and LT can be replicated by tuning the weights in proposed formulation (compare boldface entries in Tables 20.6 and 20.7 and those in Tables 20.3 and 20.5). It means that these results are all one of the noninferior solutions considering different trade-offs between mean and standard deviation (see also Figure 20.2 in the region $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$). Figure 20.3 depicts the MSE against the weight of mean response in the region $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$.

It is clear that

- LT's approach results in the lowest MSE (1634.57 in the region $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$; 2005.14 in the region $-\mathbf{1} \leq \mathbf{x} \leq \mathbf{1}$) (this corresponds to setting both weights to 0.5 in our formulation);

Table 20.4 Results of the proposed approach for target $\mathbf{T}^* = (500, 44.72)$ $(-\mathbf{1} \leq \mathbf{x} \leq \mathbf{1})$.

$\mathbf{T}^*$	ω $(\omega_\mu, \omega_\sigma)$	$\mathbf{Y}(\mathbf{X})^{\mathrm{T}}$ (y_μ, y_σ)	$\mathbf{X}^{\mathrm{T}}$ (x_1, x_2, x_3)	MSE	Delta $(\delta_\mu, \delta_\sigma)$
(500, 44.72)	(1/500, 1/44.72)	(499.9667, 45.0937)	(1,0.1183, −0.2598)	2033.44	(−16.65, 16.7119)
(500, 44.72)	(1, 1)	(499.6639, 45.0574)	(1,0.1157, −0.2593)	2030.29	(−0.3361, 0.3374)
(500, 44.72)	(500, 44.72)	(498.2014, 44.8822)	(1,0.1040, −0.2575)	2017.65	(−0.0036, 0.0036)

Table 20.5 Results of some existing methods ($-\mathbf{1} \leq \mathbf{x} \leq -\mathbf{1}$).

Methods	$Y(\mathbf{x})^T$ $(\gamma_\mu, \gamma_\sigma)$	$\mathbf{x}^T$ (x_1, x_2, x_3)	MSE	Proposed Scheme $(\omega_\mu, \omega_\sigma)$
VM and DM	(500, 45.097)	(1,0.1184, −0.259)	2033.74	(0, 1)
LT	(494.44, 44.43)	(1,0.07, −0.25)	2005.14	(0.50, 0.50)
KL	(496.08, 44.63)	(1,0.086, −0.254)	2007.07	(0.46, 0.54)

Table 20.6 Partial results of sensitivity analysis of weights ($\mathbf{x}\mathbf{x}^T \leq 3$).

ω_μ	ω_σ	y_μ	y_σ	x_1	x_2	x_3	MSE	δ_μ	δ_σ
0.00	**1.00**	**500.0000**	**40.6575**	**1.5719**	**−0.7220**	**−0.0875**	**1652.46**	**N/A**	**40.6575**
0.01	0.99	499.9982	40.6574	1.5716	−0.7230	−0.0860	1653.02	−0.1804	41.0680
…	…	…	…	…	…	…	…	…	…
0.11	**0.89**	**499.9356**	**40.6507**	**1.5717**	**−0.7228**	**−0.0866**	**1652.48**	**−0.5852**	**45.6749**
0.12	0.88	499.9195	40.6490	1.5716	−0.7228	−0.0866	1652.34	−0.6706	46.1920
…	…	…	…	…	…	…	…	…	…
0.43	0.57	497.5567	40.3977	1.5685	−0.7295	−0.0868	1637.94	−5.6822	70.8731
0.49	**0.51**	**496.0553**	**40.2382**	**1.5664**	**−0.7342**	**−0.0863**	**1634.67**	**−8.0504**	**78.8983**
0.50	**0.50**	**495.7312**	**40.2037**	**1.5659**	**−0.7352**	**−0.0862**	**1634.56**	**−8.5375**	**80.4075**
0.51	0.49	495.3807	40.1665	1.5654	−0.7363	−0.0861	1634.69	−9.0574	81.9725
…	…	…	…	…	…	…	…	…	…

Table 20.7 Partial results through sensitivity analysis of weights ($-\mathbf{1} \leq \mathbf{x} \leq \mathbf{1}$).

ω_μ	ω_σ	y_μ	y_σ	x_1	x_2	x_3	MSE	δ_μ	δ_σ
0.00	**1.00**	**500.0000**	**45.0977**	**1**	**0.1186**	**−0.2598**	**2033.80**	**N/A**	**45.0977**
0.01	0.99	499.9994	45.0977	1	0.1213	−0.2617	2033.80	−0.0551	45.5532
0.02	0.98	499.9972	45.0974	1	0.1183	−0.2596	2033.77	−0.1412	46.0177
0.03	**0.97**	**499.9948**	**45.0971**	**1**	**0.1212**	**−0.2617**	**2033.75**	**−0.1721**	**46.4919**
0.04	0.96	499.9906	45.0966	1	0.1212	−0.2617	2033.71	−0.2343	46.9757
…	…	…	…	…	…	…	…	…	…
0.46	**0.54**	**496.1257**	**44.6338**	**1**	**0.0862**	**−0.2541**	**2007.19**	**−8.4224**	**82.6552**
0.47	0.53	495.8056	44.5955	1	0.0835	−0.2536	2006.36	−8.9242	84.1425
0.48	0.52	495.4605	44.5543	1	0.0807	−0.2531	2005.69	−9.4573	85.6813
0.49	0.51	495.0879	44.5097	1	0.0776	−0.2525	2005.24	−10.0247	87.2740
0.50	**0.50**	**494.6856**	**44.4616**	**1**	**0.0743**	**−0.2519**	**2005.08**	**−10.6289**	**88.9233**
0.51	0.49	494.2509	44.4097	1	0.0707	−0.2513	2005.27	−11.2728	90.6320
…	…	…	…	…	…	…	…	…	…

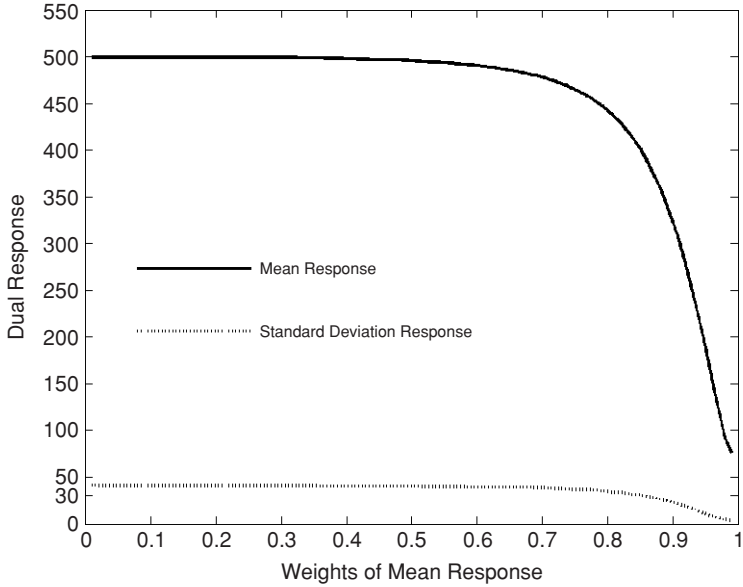


Figure 20.2 Dual response versus the weight for the mean in Example 1.

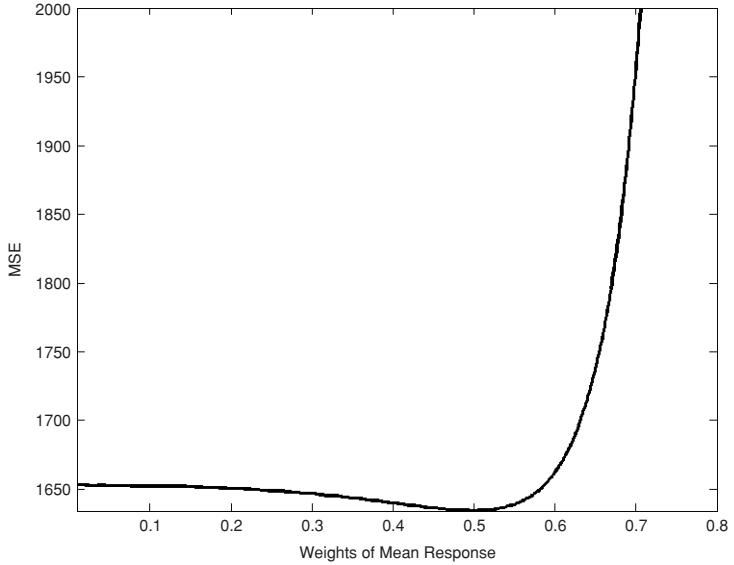


Figure 20.3 MSE versus the weight for the mean in Example 1.

- CN's approach does minimize the variance response while allowing the deviation of mean value from the target within a interval; (The result can be closely replicated by setting $\omega_\mu = 0.49$, $\omega_\sigma = 0.51$ in proposed scheme for $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$);
- KL's approach results in maximum degree of satisfaction in terms of the membership function constructed (the result can be closely replicated by setting $\omega_\mu = 0.46$, $\omega_\sigma = 0.54$ in the proposed scheme for $-\mathbf{1} \leq \mathbf{x} \leq \mathbf{1}$);
- VM and DM's approach gives the minimum variance while keeping the mean at a specified target value (500) (this corresponds to setting the weights to (0, 1) in the proposed scheme).

From the sensitivity analysis, the effect of the weights on how near $\hat{y}_\mu(\mathbf{x})$ is to T_μ^* and how near $\hat{y}_\sigma(\mathbf{x})$ is to T_σ^* can easily be inferred. For example, when the weight of the mean response increases, the deviation from the mean target increases while the deviation from the variance response target decreases. This is shown clearly in Figure 20.2. Moreover, for the problem considered, the deviation between the mean targets and the proposed solution increases rapidly when ω_μ is increased beyond 0.6. In Figure 20.3, it is observed that there is an optimal weight combination (when both weights are 0.5) that will result in minimum MSE, which is identical to the objective of LT's approach.

20.3.3 Larger-is-better and smaller-is-better

As mentioned earlier, for the larger-is-better (or smaller-is-better) case, the target for the mean should be set to the maximum (minimum) value realizable in the formulation (see Table 20.1, cases 3 and 4). These values for both mean and standard deviation are given in Table 20.8. Note that all discussions below are based on the restriction $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$ and other restrictions (radius of region of interest) can be treated by the same method.

Several results using the proposed scheme for "larger-is-better" are given in Table 20.9. The target are $T_\mu^{\max} = 952.0519$, $T_\sigma^* = 60$ (according to DM and CN). As before, our results are identical to those of DM and CN when the weights are (1, 0). (Note that their results are (672.50, 60.0) at point (1.7245, −0.0973, −0.1284) and (672.43, 60.0) at point (1.7236, −0.1097, −0.1175)). This concurs with the formulation presented in Table 20.1. The assignment of all the weight to the mean response matches our postulation that one could achieve a higher mean (or lower mean, in the next case) without optimizing the variance.

For the smaller-is-better situation, CN's result is (4.18, 35.88) at point (−0.1980, −0.2334, −1.7048) with the constraint $\hat{y}_\sigma \leq 75$. Using $T_\mu^{\min} = 3.8427$, $T_\sigma^* = 35.88$, some results of our scheme are given in Table 20.10. It may be observed that our results are at

Table 20.8 Optimal points using single objective optimization.

		Minimum	Maximum
Mean	y_μ	3.8427	952.0519
Variance	y_σ	4.0929	155.1467

Table 20.9 Partial results for 'larger-is-better'.

ω_μ	ω_σ	$\hat{y}_\mu$	$\hat{y}_\sigma$	x_1	x_2	x_3	δ_μ	δ_σ
1.00	**0.00**	**672.5048**	**60.0000**	**1.7245**	**−0.0976**	**−0.1283**	**−279.5903**	**N/A**
0.99	0.01	674.4746	60.2361	1.7250	−0.0890	−0.1285	−280.3810	23.6067
0.98	0.02	680.2988	60.9371	1.7261	−0.0648	−0.1280	−277.2991	46.8570
...	...	...	...	...	...	...	...	...
0.80	0.20	860.8103	87.8957	1.6384	0.5344	0.1736	−114.0520	139.4786
...	...	...	...	...	...	...	...	...
0.50	0.50	931.9012	107.5188	1.5124	0.6717	0.5112	−40.3014	95.0376
...	...	...	...	...	...	...	...	...
0.01	0.99	952.0052	123.1573	1.3615	0.7298	0.7835	−4.6699	63.7953

least as good as CN's results; in particular, the result (3.8435, 36.4030) is better than that of CN since the value of mean 3.8435 is smaller than 4.18 with the value of standard deviation (36.4030) is within the region $\hat{y}_\sigma \leq 75$.

20.4 EXAMPLE 2

The experiment in Example 1 involved a 3^3 design with each setting run three times to give a total of 81 runs.[18] From this design, Vining and Schaub[24] extracted a partially replicated factorial, face-centered central composite design. The resulting design consisted of 22 runs and the dual response surfaces in which a linear variance model is fitted are as follows:

$$\hat{y}_\mu = 304.8 + 162.9x_1 + 102.3x_2 + 148.4x_3 - 16.7x_1^2 + 12.7x_2^2 + 46.8x_3^2 + 79.1x_1x_2 + 77.5x_1x_3 + 55.2x_2x_3,$$

$$\hat{y}_\sigma = 86.7 + 6.0x_1 - 1.6x_2 + 51.5x_3.$$

Table 20.10 Partial results for 'smaller-is-better'.

ω_μ	ω_σ	y_μ	y_σ	x_1	x_2	x_3	δ_μ	δ_σ
1.00	**0.00**	**4.1807**	**35.8800**	**−0.1932**	**−0.2306**	**−1.7057**	**0.3381**	**N/A**
0.99	0.01	4.1809	35.8798	−0.1941	−0.2311	−1.7055	0.3416	−0.0164
0.98	0.02	4.1806	35.8801	−0.1939	−0.2311	−1.7056	0.3448	0.0064
...	...	...	...	...	...	...	...	...
0.80	0.20	4.1647	35.8979	−0.1921	−0.2334	−1.7055	0.4025	0.0896
...	...	...	...	...	...	...	...	...
0.50	0.50	4.0445	36.0404	−0.1822	−0.2568	−1.7032	0.4037	0.3208
...	...	...	...	...	...	...	...	...
0.01	**0.99**	**3.8435**	**36.4030**	**−0.1459**	**−0.3396**	**−1.6922**	**0.0833**	**0.5283**

Table 20.11 Partial results of sensitivity analysis of weights ($\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$).

ω_μ	ω_σ	y_μ	y_σ	x_1	x_2	x_3	MSE	δ_μ	δ_σ
0.00	**1.00**	**500**	**52.9978**	**0.9498**	**1.2532**	**−0.7261**	**2808.765**	**N/A**	**52.9978**
0.01	**0.99**	**499.9991**	**52.9990**	**0.9563**	**1.2478**	**−0.7270**	**2808.892**	**−0.0904**	**53.5343**
...	...	...							
0.04	0.96	499.9855	52.9955	0.9501	1.2530	−0.7262	2808.524	−0.3621	55.2036
0.05	0.95	499.9769	52.9942	0.9501	1.2530	−0.7263	2808.38	−0.4622	55.7833
...	...	...	...	...	...	...	...	...	...
0.49	0.51	492.4574	51.8094	0.9417	1.2462	−0.7485	2741.104	−15.393	101.5871
0.50	**0.50**	**491.8457**	**51.7129**	**0.9414**	**1.2453**	**−0.7504**	**2740.721**	**−16.3086**	**103.4259**
0.51	0.49	491.1807	51.6081	0.9404	1.2448	−0.7523	2741.171	−17.2927	105.3226
...	...	...	...	...	...	...	...	...	...

20.4.1 Target-is-best

Similarly, we use the convex combination of weights in our scheme to investigate the target-is-best case (where one seeks to minimize $\hat{y}_\sigma$ subject to $\hat{y}_\mu = 500$) in the region of $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$. Using $T_\mu^* = 500$, $T_\sigma^* = 0$, some results of our scheme are given in Table 20.11.

The recommended setting for $\mathbf{x}\mathbf{x}^{\mathrm{T}} = 3$ by Vining and Schaub[24] is (0.993, 1.216, −0.731) with $\hat{y}_\sigma = 56.95$; from Table 20.11, it is clear that our model results in the setting (0.9498, 1.2532, −0.7261) with $\hat{y}_\sigma = 52.9978$ and (0.956, 1.247, −0.727) with $\hat{y}_\sigma = 52.999$ using the smaller design of Vining and Schaub[23] (22 runs). The result (when $\omega_\mu = 0.01$, $\omega_\sigma = 0.99$) is very close to that recommended by Del Castillo and Montgomery,[3] (0.953, 1.246, −0.735), for the more complex design (81 runs). This highlights the advantage of having the proposed tuning scheme in which the trade-off between dual responses can be explicitly considered. This trade-off between dual responses against the weight of mean response is depicted in Figure 20.4. Figure 20.5 depicts the MSE against the weight of mean response in the region $\mathbf{x}\mathbf{x}^{\mathrm{T}} \leq 3$. These figures result in the same conclusions as those in Example 1, and thus further support our approach.

20.5 CONCLUSIONS

In this chapter, a unified approach for optimizing the dual response system has been presented. The proposed approach allows for explicit consideration of the trade-off between meeting the mean and variance targets. This is not considered in other techniques. Moreover, the solutions of our scheme provide not only the slackness from the mean target, but also the slackness from the variance response target which is not available from other formulations. It was shown that some of the existing formulations can be represented using our approach. In particular, through two classical examples, we showed that the other techniques can be modeled or closely modeled by the proposed scheme through a parametric adjustment scheme, and that the results from other existing techniques are noninferior solutions. As our formulation can easily be implemented using Excel Solver as well as other software packages, the approach is both general and practical.

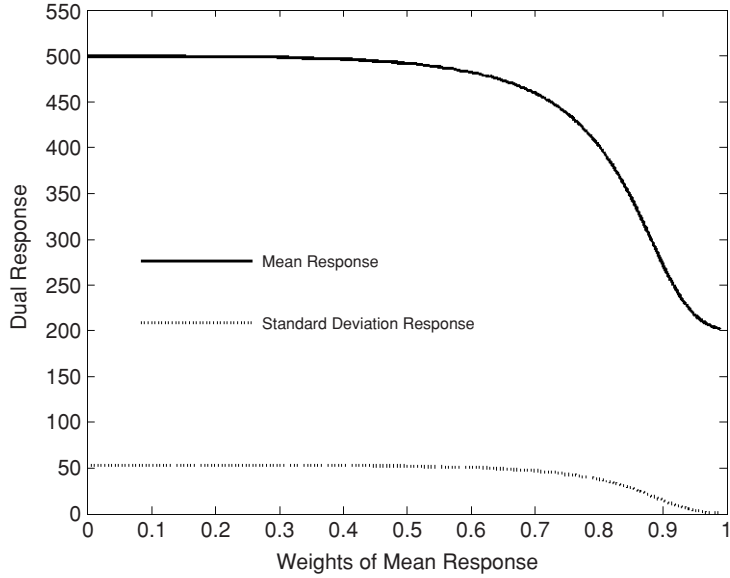


Figure 20.4 Dual response versus the weight for the mean in Example 2.

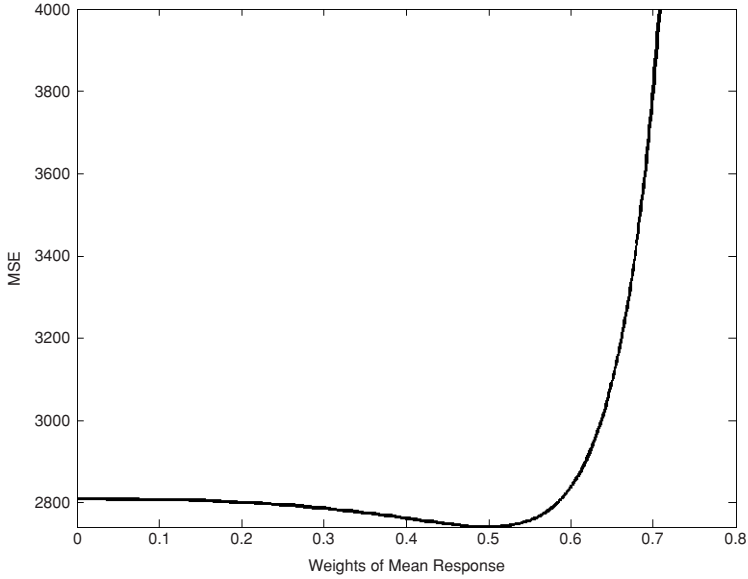


Figure 20.5 MSE versus the weight for the mean in Example 2.

REFERENCES

1. Myers, R.H. and Carter, W.H. (1973) Response surface techniques for dual response systems. *Technometrics*, 15, 301–317.
2. Vining, G.G. and Myers, R.H. (1990) Combining Taguchi and response surface philosophies: a dual response approach. *Journal of Quality Technology*, 22, 38–45.
3. Del Castillo, E. and Montgomery, D.C. (1993) A nonlinear programming solution to the dual response problem. *Journal of Quality Technology*, 25, 199–204.
4. Lin, D.K.J. and Tu, W. (1995) Dual response surface optimization. *Journal of Quality Technology*, 27, 34–39.
5. Copeland, K.A.F. and Nelson, P.R. (1996). Dual response optimization via direct function minimization. *Journal of Quality Technology*, 28, 331–336.
6. Ames, A.E., Mattucci, N., Macdonald, S., Szonyi, G. and Hawkins, D.M. (1997) Quality loss functions for optimization across multiple response surfaces. *Journal of Quality Technology,* 29, 339–346.
7. Kim, K.K., and Lin, D.K.J. (1998) Dual response surface optimization: A fuzzy modeling approach. *Journal of Quality Technology*, 30, 1–10.
8. Derringer, G. and Suich, R. (1980) Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12, 214–219.
9. Vining, G.G. (1998) A compromise approach to multiple response optimization. *Journal of Quality Technology*, 30, 309–313.
10. Khuri, A.I. and Conlon, M. (1981) Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics*, 23, 363–375.
11. Pignatielio Jr., J.J. (1993) Strategies for robust multiresponse quality engineering. *IIE Transactions*, 25, 5–15.
12. Ignizio, J.P. (1976). *Goal Programming and Extensions.* Lexington, MA: Lexington Books.
13. Taha, H.A. (1997) *Operations Research*, 6th edition. Upper Saddle River, NJ: Prentice Hall.
14. Censor Y. (1977) Pareto optimality in multiobjective problems. *Applied Mathematics and Optimization*, 4, 41–59.
15. Del Castillo, E., Fan, S.K. and Semple, J. (1997) The computation of global optima in dual response systems. *Journal of Quality Technology,* 29, 347–353.
16. Del Castillo, E., Fan, S.K. and Semple J. (1999) Optimization of dual response systems: a comprehensive procedure for degenerate and nondegenerate problems. *European Journal of Operationals Research*, 112, 174–186.
17. Fan, S.K. (2000) A generalized global optimization algorithm for dual response systems. *Journal of Quality Technology*, 32, 444–456.
18. Box, G.E.P and Draper, N.R. (1987) *Empirical Model Building and Response Surfaces*. New York: John Wiley & Sons, Inc.
19. Myers, R.H. and Montgomery, D.C. (1995) *Response Surface Methodology*. New York: John Wiley & Sons, Inc.
20. Khuri, A. and Cornell, J. (1996) *Response Surfaces: Designs and Analysis*, New York: Dekker.
21. Myers, R.H. (1999) Response surface methodology – current status and future directions. *Journal of Quality Technology* 31, 30–44.
22. Cohon, J.L. (1978) *Multiobjective Programming and Planning*. New York: Academic Press.
23. Ehrgott, M. (2000) *Multicriteria Optimization*. New York, Springer-Verlag.
24. Vining, G.G. and Schaub, D. (1996) Experimental designs for estimating both mean and variance function. *Journal of Quality Technology*, 28, 135–147.

Part E

Control Phase

21

Establishing Cumulative Conformance Count Charts

L. C. Tang and W. T. Cheong

The cumulative conformance count (CCC) chart has been used for monitoring high yield processes with very low process fraction nonconforming. Current work has yet to provide a systematic treatment for establishing the chart, particularly when the process fraction nonconforming parameter is estimated. We extend the results from the recent studies by Tang and Cheong and by Yang *et al.* to enable engineers to construct the CCC chart under different sampling and estimation conditions. We first outline the statistical properties of the CCC chart. We then give new insights into the behavior of CCC chart when the parameter is estimated. We propose some procedures for constructing the CCC chart when the process fraction nonconforming is given, when it is estimated sequentially, and when it is estimated with a fixed sample size. The proposed steps are implemented using data from a high-yield process in order to demonstrate the effectiveness of the scheme.

21.1 INTRODUCTION

Attribute Shewhart charts (such as the p chart) have proven their usefulness, but are ineffective as the level of nonconformance improves to low levels.[1] This has led to the development of new methods of process monitoring, one of which is the cumulative conformance count (CCC) chart. CCC charts track the number of conforming items produced between successive nonconforming ones. This type of chart has been shown

This chapter is based on the article by L. C. Tang and W. T. Cheong, "On establishing cumulative conformance charts", International Journal of Performability Engineering, 1(1), 2005, pp. 5–22, and is reproduced by the permission of the publisher, the RAMS Consultants.

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

to be very useful in monitoring p, the process fraction nonconforming, for high-quality processes. Most studies of CCC charts assume that p is known.[1–4] However, in practice, p is not likely to be known before implementing the chart and will need to be estimated. Though Tang and Cheong[5] and Yang *et al.*[6] examine the effects of parameter estimation on the CCC chart, a systematic framework for establishing the CCC chart under different ways of estimating p has not been presented. For example, to implement the CCC chart using Tang and Cheong's scheme,[5] one must know when to stop updating the estimate of p so that the control limits are not affected by data from processes with drift. If Yang *et al.*'s scheme[6] is used, the initial sample size must be specified.

In the following, the statistical properties of CCC charts and some preliminary results on the effects of parameter estimation are revisited. This is followed by guidelines for establishing CCC charts for schemes proposed by both Tang and Cheong[5] and Yang *et al.*[6] Numerical examples for designing CCC charts are given for both cases when the process fraction nonconforming, p, is given or unknown to illustrate the applicability of the proposed guidelines.

21.2 BASIC PROPERTIES OF THE CCC CHART

The CCC chart tracks the number of conforming items produced before a nonconforming one is produced. As a result, the probability of the nth item being the first nonconforming item to be discovered is given by

$$P\{X = n\} = (1 - p)^{n-1} p, \qquad n = 1, 2, \ldots, \tag{21.1}$$

which is a geometric distribution with parameter p. It is assumed that inspections are carried out sequentially and the above random variables are independently and identically distributed.

In CCC charts, the control limits are determined based on the probability limits from the geometric model shown in equation (21.1). For a given probability of type I error, α, the two-sided control limits are given by

$$UCL = \frac{\ln \alpha/2}{\ln (1 - p_0)}, \tag{21.2}$$

$$LCL = \frac{\ln (1 - \alpha/2)}{\ln (1 - p_0)} + 1, \tag{21.3}$$

where p_0 is the in-control fraction nonconforming.[5] The resulting average run length (ARL), from the above limits, initially increases when the process starts to deteriorate (p increases) and decreases after attaining a maximum point at $p > p_0$. This is a common problem for data having a skewed distribution,[7] and renders the CCC chart rather insensitive in detecting an increase in p and may lead to the misinterpretation that the process is well in control, or has been improved.

To overcome the problem mentioned above, Xie *et al.*[7] showed that, for a given p_0 and initial type I error rate ϕ, an adjustment factor, γ_ϕ, given by

$$\gamma_\phi = \frac{\ln [\ln(1 - \phi/2)/\ln(\phi/2)]}{\ln [(\phi/2)/\ln(1 - \phi/2)]} \tag{21.4}$$

can be applied to the control limits; that is,

$$UCL_{\gamma_\phi} = \frac{\ln(\alpha/2)}{\ln(1-p_0)}\gamma_\phi, \tag{21.5}$$

$$LCL_{\gamma_\phi} = \left(\frac{\ln(1-\alpha/2)}{\ln(1-p_0)}\right)\gamma_\phi + 1, \tag{21.6}$$

so that the ARL is maximized at $p = p_0$, which is the in-control ARL, ARL_0. It should be noted that the above control limits do not have a direct probability interpretation; that is, they do not always give the same probability content within the control limits. Notably, even though the control limits are shifted by a factor of γ_ϕ, the type I risk of the chart is no longer given by the initial value ϕ. Moreover, the false alarm probabilities beyond both sides of the control limits are no longer the same. The adjusted control limits with equal probabilities are considered by Zhang *et al.*[8] However, their proposed procedure involves several iterations and only considers known p.

The actual type I risk α_0 can be recomputed by summing the tail probabilities $\alpha_u = P\left\{X > UCL_{\gamma_\phi}\right\}$ and $\alpha_l = P\left\{X < LCL_{\gamma_\phi}\right\}$, where X is the number of conforming items between two adjacent nonconforming items. From equations (21.5) and (21.6),

$$\alpha_u = (1-p_0)^{\ln(\alpha/2)\gamma_\phi/\ln(1-p_0)} \quad \text{and} \quad \alpha_l = 1-(1-p_0)^{\ln(1-\alpha/2)\gamma_\phi/\ln(1-p_0)},$$

which can be expressed as

$$\alpha_u = (\alpha/2)^{\gamma_\phi} \quad \text{and} \quad \alpha_l = 1-(1-\phi/2)^{\gamma_\phi}, \tag{21.7}$$

respectively. It follows that

$$\alpha_0 = \alpha_u + \alpha_l = (\phi/2)^{\gamma_\phi} - (1-\phi/2)^{\gamma_\phi} + 1. \tag{21.8}$$

As a result, α_0 is a nonlinear function of ϕ, and, despite the apparent relation between ϕ and p_0 in equations (21.5) and (21.6), they are in fact independent of each other.

21.3 CCC SCHEME WITH ESTIMATED PARAMETER

In most circumstances, p_0 is not likely to be known and needs to be estimated. Tang and Cheong[5] presented a sequential estimation scheme, while Yang *et al.*[6] considered using the conventional binomial estimator of p for the CCC scheme.

21.3.1 Sequential estimation scheme

Tang and Cheong[5] use the unbiased estimator of p given by Haldane,[9]

$$\bar{p} = \frac{m-1}{N_m - 1}, \tag{21.9}$$

where the number of nonconforming items to be observed, m, is fixed *a priori* and the total number of samples to be inspected, N_m, is a random variable.

In the scheme proposed by Tang and Cheong,[5] $\bar{p}$ is updated sequentially and the control limits are revised so that the ARL_0 of the chart not only can be kept to the desired value, but also is the peak of the ARL curve. This can be done by making use of the same adjustment given in equations (21.5) and (21.6). The control limits are

$$UCL = \frac{\ln(\phi_m/2)}{\ln(1-(m-1)/(N_m-1))} \times \gamma_{\phi_m}, \tag{21.10}$$

$$LCL = \left(\frac{\ln(1-\phi_m/2)}{\ln(1-(m-1)/(N_m-1))}\right)\gamma_{\phi_m} + 1, \tag{21.11}$$

where ϕ_m can be obtained by solving

$$\sum_{n=m}^{\infty}\left((1-\bar{p})^{\ln(\phi_m/2)\gamma_{\phi_m}/\ln(1-\bar{p})} - (1-\bar{p})^{\ln(1-\phi_m/2)\gamma_{\phi_m}/\ln(1-\bar{p})} + 1\right)^{-1} \times \binom{n-1}{m-1}\bar{p}^m(1-\bar{p})^{n-m} = ARL_0, \tag{21.12}$$

with $\bar{p}$ given by equation (21.9) and γ_{ϕ_m} given by (21.4).

Using similar steps to those proceeding (21.8), ϕ_m is shown to be independent of $\bar{p}$. Tang and Cheong[5] showed that, as m increases, the sensitivity of the chart improves and approaches that with known process parameter. Here, the index ρ, reflecting the shift in p, is used; that is, $p = \rho p_0$. When $\rho = 1$, the process is in statistical control; if $\rho > 1$, the process is deteriorating; and $\rho < 1$ indicates process improvement.

To obtain the ARL curves for different m using the proposed scheme, equation (21.12) can be re-written as

$$\sum_{n=m}^{\infty}\left((1-\bar{p})^{\ln(\phi_m/2)\gamma_{\phi_m}/\ln(1-\bar{p})} - (1-\bar{p})^{\ln(1-\phi_m/2)\gamma_{\phi_m}/\ln(1-\bar{p})} + 1\right)^{-1} \times \binom{n-1}{m-1}\bar{p}^m(1-\bar{p})^{n-m} = ARL_m, \tag{21.13}$$

where ARL_m is the ARL with $\bar{p}$ estimated after observing m nonconforming items. Figure 21.1 depicts the ARL curves of the CCC scheme under sequential estimation, using $m = 5$, 10, 30 and 50, for $ARL_0 = 370$ and $p_0 = 500$ ppm.

21.3.2 Conventional estimation scheme

On the other hand, Yang *et al.*[6] considered using the conventional estimator,

$$\hat{p} = \frac{D_n}{n}, \tag{21.14}$$

where n is the initial sample size, fixed a priori, and D_n is the number of nonconforming items among n items sampled.

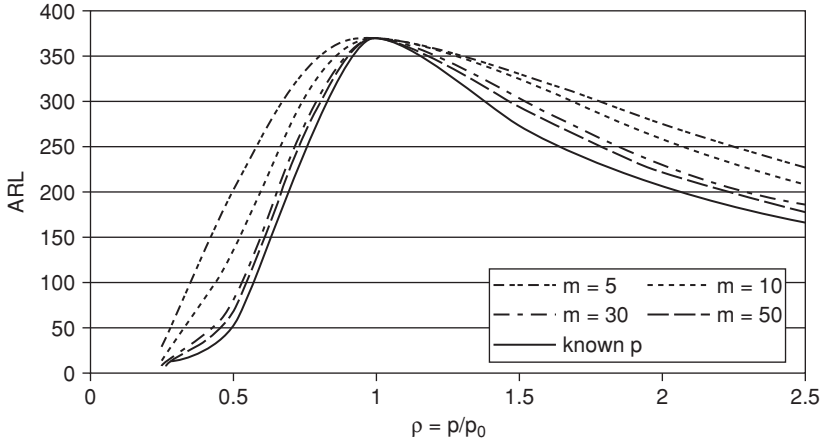


Figure 21.1 ARL under sequential estimation with $m = 5$, 10, 30 and 50, given $ARL_0 = 370$ and $p_0 = 500$ ppm.

Yang *et al.*[6] investigated the sample size effect and gave the exact false alarm probability equation when p_0 is estimated using the conventional estimator. Using this equation, a similar adjustment scheme proposed by Tang and Cheong[5] can be applied so that the ARL_0 is the maximum point of the ARL curve. Thus, the equation becomes

$$\sum_{d=0}^{n} \left((1-\hat{p})^{\ln(\phi_n/2)\gamma_{\phi_n}/\ln(1-\hat{p})} - (1-\hat{p})^{\ln(1-\phi_n/2)\gamma_{\phi_n}/\ln(1-\hat{p})} + 1 \right)^{-1} \times \binom{n}{d} \hat{p}^d (1-\hat{p})^{n-d} = ARL_0 \tag{21.15}$$

and the control limits become

$$UCL = \frac{\ln(\phi_n/2)}{\ln(1-(D_n/n))}\gamma_{\phi_n}, \tag{21.16}$$

$$LCL = \left(\frac{\ln(1-\phi_n/2)}{\ln(1-(D_n/n))} \right) \gamma_{\phi_n} + 1, \tag{21.17}$$

where ϕ_n can be obtained from solving (21.15), with $\hat{p}$ given in (21.14) and γ_{ϕ_n} from equation (21.4), after specifying the desired ARL_0.

From the study of Yang *et al.*,[6] the larger the sample size used in estimating p_0, the closer the chart performs to that with known parameter. Thus, it is expected that the performance of the scheme proposed here approaches that with known p_0, as the sample size used to estimate p_0 increases. For illustration, Figure 21.2 shows the ARL curves for known p_0, $n = 10\,000$, $20\,000$, $50\,000$, and $100\,000$, using the proposed conventional estimator scheme, with ARL_0 set at 370 and $\hat{p} = 0.0005$.

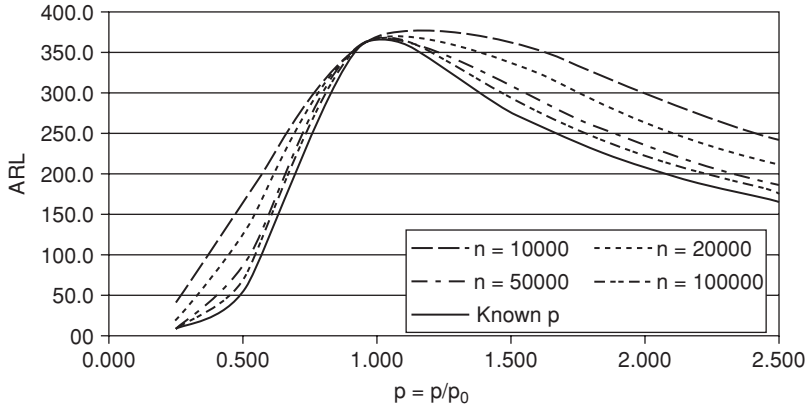


Figure 21.2 ARL for known p (500 ppm), $n = 10\,000$, 20 000, 50 000, and 100 000, using the conventional estimator with ARL_0 set at 370 and $\hat{p} = 0.0005$.

21.4 CONSTRUCTING A CCC CHART

When monitoring a process with given p_0, the control limits can easily be computed from equations (21.5) and (21.6).

The ARL_0 is the average run length of the CCC chart while the process is in statistical control; it is the reciprocal of α_0, and is independent of p_0. After specifying the preferred ARL_0, the parameter ϕ can be obtained by solving

$$\frac{1}{(\phi/2)^{\gamma_\phi} - (1-\phi/2)^{\gamma_\phi} + 1} = ARL_0, \tag{21.18}$$

where γ_ϕ is a function of ϕ given by equation (21.4). The corresponding control limits can then be computed. Table 21.1 gives the values of ϕ and γ_ϕ with different ARL_0, which can be substituted into equations (21.5) and (21.6) to determine the control limits.

21.4.1 Establishing a CCC chart with the sequential estimator

In an industrial setting where inspections are carried out sequentially, parameter estimation as well as process monitoring can be started once there are two nonconforming

Table 21.1 The values of ϕ and respective adjustment factor γ_ϕ with different ARL_0.

ARL_0	ϕ	γ_ϕ
200	0.006 75	1.306 03
370	0.003 73	1.292 69
500	0.002 78	1.286 54
750	0.001 88	1.278 63
1000	0.001 42	1.273 28

Table 21.2 The values of ϕ_m and γ_ϕ for different m and preferred ARL_0 for a CCC scheme with sequential sampling plan.

ARL	200		370		500		750		1000	
m	ϕ_m	γ_{Φ_m}	ϕ_m	γ_{Φ_m}	ϕ_m	γ_{Φ_m}	ϕ_m	γ_{Φ_m}	ϕ_m	γ_{Φ_m}
2	0.003 63	1.2921	0.001 96	1.2795	0.001 45	1.2737	0.000 97	1.2663	0.000 73	1.2613
3	0.004 24	1.2955	0.002 29	1.2826	0.001 69	1.2767	0.001 13	1.2691	0.000 85	1.2640
4	0.004 59	1.2972	0.002 48	1.2842	0.001 83	1.2782	0.001 22	1.2705	0.000 92	1.2654
5	0.004 83	1.2984	0.002 61	1.2852	0.001 93	1.2792	0.001 29	1.2715	0.000 97	1.2663
6	0.005 01	1.2992	0.002 71	1.2860	0.002 01	1.2800	0.001 34	1.2722	0.001 00	1.2669
7	0.005 16	1.2998	0.002 79	1.2866	0.002 07	1.2805	0.001 38	1.2728	0.001 03	1.2675
8	0.005 28	1.3004	0.002 86	1.2871	0.002 12	1.2810	0.001 41	1.2732	0.001 06	1.2679
9	0.00538	1.3008	0.002 92	1.2875	0.002 16	1.2814	0.001 44	1.2736	0.001 08	1.2683
10	0.005 47	1.3012	0.002 97	1.2879	0.002 20	1.2818	0.001 47	1.2739	0.001 10	1.2686
20	0.005 95	1.3031	0.003 25	1.2897	0.002 41	1.2836	0.001 61	1.2757	0.001 21	1.2704
30	0.006 17	1.3039	0.003 37	1.2906	0.002 51	1.2844	0.001 68	1.2765	0.001 27	1.2712
50	0.006 38	1.3047	0.003 49	1.2913	0.002 60	1.2852	0.001 75	1.2773	0.001 32	1.2719
70	0.006 47	1.3050	0.003 55	1.2917	0.002 65	1.2855	0.001 78	1.2776	0.001 34	1.2723
100	0.006 55	1.3053	0.003 60	1.2920	0.002 69	1.2858	0.001 81	1.2779	0.001 36	1.2725
∞	**0.006 75**	**1.3060**	**0.003 73**	**1.2927**	**0.002 78**	**1.2865**	**0.001 88**	**1.2786**	**0.001 42**	**1.2733**

items observed ($m = 2$). Control limits can be calculated from equations (21.10) and (21.11) with ϕ_m based on the required ARL_0 from equation (21.12). The adjustment factor γ_{ϕ_m} can be obtained from (21.4).

Table 21.2 gives the values of ϕ_m and the respective adjustment factor γ_{ϕ_m} for different values of m ranging from 2 to 100 and $ARL_0 = 200, 370, 500, 750, 1000$. The last row of the table ($m = \infty$) is the value when p_0 is given, which is ϕ from Table 21.1. Similarly, Tang and Cheong[5] noted that when p is estimated sequentially, it has no effect on ϕ_m, and ϕ_m converges to that of the known value chart as m increases.

To start using the CCC chart without undue delay, it was recommended by Tang and Cheong[5] that p_0 be estimated using $m = 2$ and be sequentially updated as long as the process is deemed to be in control. Two associated problems are that the ARL performance in detecting process shift is poor for small m, and that excessive updating increases the risk of including data from processes with drift. To mitigate these problems, we propose a two-pronged approach. First, a target ARL performance for a given change in p is specified, so that the updating is terminated at a particular m; and more stringent criteria are used to decide whether updating should be suspended.

21.4.1.1 Termination of sequential updating

For a given change in p, which can be indexed by $\rho = p/p_0$, define R_s as the ratio between a desired ARL, ARL_m, and the ARL with known parameter ARL_∞. In other words,

$$R_s = \frac{ARL_m}{ARL_\infty}, \tag{21.19}$$

where the values of ARL_m can be calculated from equation (21.13) with respective values of ρ and m, and ARL_∞ is the out-of-control ARL with respect to ρ for a given p_0, given by

$$ARL_\infty = \frac{1}{(1-\rho p_0)^{UCL} + 1 - (1-\rho p_0)^{LCL-1}}. \tag{21.20}$$

The choice of R_s reflects the user's desirability of having a CCC chart that performs close to the ideal case. When the ratio $R_s \to 1$, the performance of the scheme is comparable with that of a known parameter CCC chart.

For a given ρ and ARL_0, Table 21.3 gives the corresponding m required, m^*, in achieving the desired R_s. For example, if the user wishes to establish a CCC chart with ARL_0 set at 200, and the required ARL performance at $\rho \geq 2$ is given by $R_s = 1.150$, then, from Table 21.3, one should keep updating the estimate of p and the control limits up to $m = 19$, as long as the suspension rule described below is not violated.

21.4.1.2 Suspension of sequential update

The updating must be done judiciously to avoid using the contaminated data from processes with drift which will affect the sensitivity of the control scheme. To mitigate this problem, a more stringent criterion than the usual out-of-control rule is imposed for the suspension of sequential estimation.

Two decision lines corresponding to the 5th and 95th percentile points are used to create a warning zone. Figure 21.3 shows the warning zones of the CCC chart (the unshaded areas in the chart). The center line of the graph is the 50th percentile of the cumulative distribution function (CDF) and is given by

$$CL = \frac{\ln 0.5}{\ln(1-\bar{p})} + 1 = -\frac{0.693}{\ln(1-\bar{p})} + 1, \tag{21.21}$$

and the two decision lines are given by

$$UDL = \frac{\ln(0.05)}{\ln(1-\bar{p})} \quad \text{and} \quad LDL = \frac{\ln(1-0.05)}{\ln(1-\bar{p})} + 1.$$

The updating is suspended after observing any one of the following:

1. One point beyond the shaded area, which will occur with probability 0.05.
2. Four consecutive points on one side of the center line, which will occur with probability 0.0625.

These simple rules provide adequate protection against using contaminated data for updating. Following the suspension, process monitoring should be continued without revising the control limits. A new sequential estimate of p is then initiated until the same value of m before suspension occurs. If there is no significant difference between the two estimates, all data can then be combined to give the estimate of p and updating is continued until $m = m^*$ provided that the process remains in control. If another suspension criterion is observed during this period, the same decision process is repeated.

Table 21.3 The values of m^* for different ρ and $ARL_0 = 200, 370, 500, 750$ and 1000.

	$\text{ARL}_0 = 200$					
ARL_m/ARL_∞	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1.5$	$\rho = 2.0$	$\rho = 2.5$
1.100	50	105	29	25	27	29
1.125	48	84	21	18	23	23
1.150	40	71	17	14	19	19
1.175	30	61	14	6	16	16
1.200	27	61	10	2	13	14
1.225	24	49	9	2	10	11
1.250	22	44	7	2	8	9
	$\text{ARL}_0 = 370$					
1.100	55	126	40	29	35	35
1.125	45	104	28	20	26	27
1.150	39	87	22	17	20	20
1.175	33	75	18	10	17	18
1.200	30	66	16	2	15	16
1.225	27	60	14	2	12	13
1.250	25	54	10	2	10	11
	$\text{ARL}_0 = 500$					
1.100	58	130	43	30	34	36
1.125	48	110	31	23	27	27
1.150	40	93	25	18	21	22
1.175	35	81	20	14	18	19
1.200	31	72	17	7	16	16
1.225	28	64	15	6	13	14
1.250	26	59	13	6	10	12
	$\text{ARL}_0 = 750$					
1.100	63	141	49	35	37	13
1.125	51	123	40	27	30	15
1.150	44	106	30	20	24	17
1.175	38	92	25	17	19	20
1.200	34	81	21	10	17	24
1.225	33	73	18	7	14	30
1.250	28	66	16	7	12	40
	$\text{ARL}_0 = 1000$					
1.100	66	145	55	36	39	40
1.125	54	131	44	26	30	30
1.150	46	115	32	21	25	25
1.175	40	99	28	17	20	21
1.200	36	88	23	14	18	18
1.225	32	78	20	8	15	16
1.250	30	71	18	8	14	14

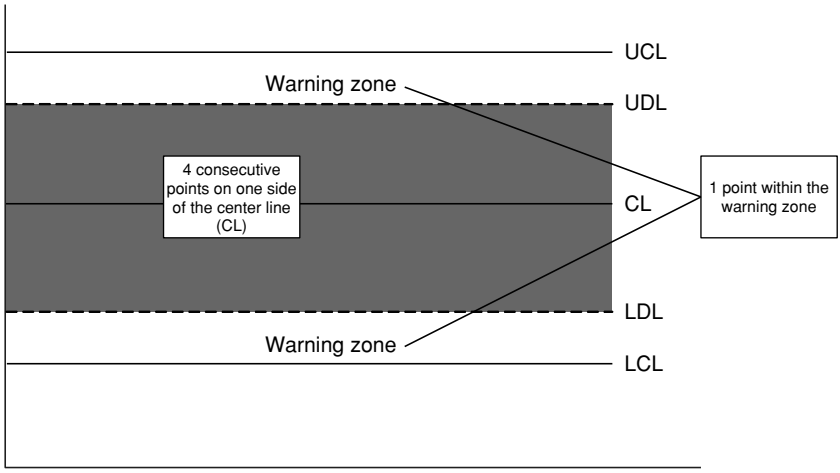


Figure 21.3 Warning zones o the CCC chart.

21.4.2 Establishing a CCC chart with the conventional estimator

When the conventional estimator of p is used, the control limits of the CCC chart can be obtained from equations (21.16) and (21.17) after estimating p_0 from the initial sample. Similarly, to achieve the desired ARL_0, ϕ_n and γ_n can be obtained from (21.15) and (21.4) respectively, given the sample size n and $\hat{p}$.

To facilitate the construction of the CCC chart, Table 21.4 gives the values of ϕ_n for different n and $\hat{p}$ ranging from 0.0001 to 0.001, with $ARL_0 = 370$, when p_0 is estimated using the conventional estimator. The last row of the table ($n = \infty$) is the value where p_0 is given, which is ϕ from Table 1. As expected, the value of ϕ_n approaches ϕ as the sample size increases. The values given in this table can be used as the input for constructing the CCC chart if the desired ARL_0 is 370. Unlike Tables 21.1 and 21.2, these ϕ_n values are dependent on $\hat{p}$.

It also worth noting that in adopting the conventional estimator there could be no nonconforming items in the initial sample. This will lead to a situation where sample size is increased incrementally until some arbitrary numbers of nonconforming items are observed. In doing so, the resulting estimate of p_0 will be biased. A simple way of avoiding this problem is to ensure that the probability of having at least one nonconforming item is sufficiently large in the initial sample. For example, the sample size for a preliminary value of $p_0 = 100$ ppm and a 90% chance of observing at least one nonconforming item is

$$n = \frac{\ln(0.1)}{\ln(1 - p_0)} \approx 23\,000.$$

Nevertheless, Yang *et al.*[6] concluded that the sample size used for estimation should be large enough for better performance of the chart, which is evident from Figure 21.2. An updating scheme similar to that of the sequential estimate can be adopted.

Table 21.4 The values of ϕ_n for different n and $\hat{p}$ with $ARL_0 = 370$ for a CCC scheme using a binomial sampling plan.

	$\hat{p}$									
n	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009	0.001
10 000	0.001 48	0.001 92	0.002 22	0.002 42	0.002 57	0.002 68	0.002 77	0.002 84	0.002 91	0.002 96
20 000	0.001 92	0.002 42	0.002 68	0.002 84	0.002 96	0.003 05	0.003 11	0.003 17	0.003 21	0.003 25
50 000	0.002 57	0.002 96	0.003 14	0.003 25	0.003 33	0.003 38	0.003 42	0.003 45	0.003 48	0.003 50
100 000	0.002 96	0.003 25	0.003 38	0.003 45	0.003 50	0.003 53	0.003 56	0.003 58	0.003 59	0.003 60
200 000	0.003 25	0.003 45	0.003 53	0.003 58	0.003 60	0.003 62	0.003 64	0.003 64	0.003 66	0.003 66
300 000	0.003 38	0.003 53	0.003 59	0.003 62	0.003 64	0.003 66	0.003 67	0.003 67	0.003 68	0.003 68
400 000	0.003 45	0.003 58	0.003 62	0.003 65	0.003 66	0.003 67	0.003 68	0.003 69	0.003 69	0.003 69
500 000	0.003 50	0.003 60	0.003 64	0.003 66	0.003 67	0.003 68	0.003 69	0.003 69	0.003 70	0.003 70
600 000	0.003 53	0.003 62	0.003 66	0.003 67	0.003 68	0.003 69	0.003 69	0.003 70	0.003 70	0.003 70
700 000	0.003 56	0.003 64	0.003 67	0.003 68	0.003 69	0.003 69	0.003 70	0.003 70	0.003 70	0.003 71
800 000	0.003 58	0.003 65	0.003 67	0.003 69	0.003 69	0.003 70	0.003 70	0.003 70	0.003 71	0.003 71
900 000	0.003 59	0.003 66	0.003 68	0.003 69	0.003 70	0.003 70	0.003 70	0.003 71	0.003 71	0.003 71
1 000 000	0.003 60	0.003 66	0.003 68	0.003 69	0.003 70	0.003 70	0.003 71	0.003 71	0.003 71	0.003 71
2 000 000	0.003 66	0.003 69	0.003 69	0.003 71	0.003 71	0.003 71	0.003 72	0.003 72	0.003 72	0.003 72
∞	0.003 73	0.003 73	0.003 73	0.003 73	0.003 73	0.003 73	0.003 73	0.003 73	0.003 73	0.003 73

Table 21.5 Values of R_c for different ρ and n, with $ARL_0 = 370$ for $\hat{p} = 0.0005$.

n	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1.5$	$\rho = 2.0$	$\rho = 2.5$
10 000	5.446	3.021	1.198	1.319	1.433	1.455
20 000	2.529	2.323	1.156	1.217	1.260	1.264
50 000	1.293	1.584	1.098	1.112	1.122	1.122
100 000	1.118	1.279	1.064	1.063	1.067	1.067
1 000 000	1.011	1.025	1.010	1.008	1.008	1.008

Define the ARL ratio, R_c as

$$R_c = \frac{ARL_n}{ARL_\infty}, \tag{21.22}$$

where ARL_n is the ARL in detecting process shift by a factor of ρ, when a total sample of size n is used for estimating p, and the ARL_∞ is the ARL with known parameter.

Once ARL_0 is specified, for a given $\hat{p}$ (from initial estimate) and a process shift of interest indexed by ρ, the minimum sample size needed, n^*, to achieve a certain R_c requirement can be determined. Table 21.5 gives the values of R_c for different ρ and n, with $ARL_0 = 370$ for $\hat{p} = 0.0005$. For example, given the estimated value $\hat{p} = 0.0005$, for ARL_0 specified at 370, and assuming that the required ARL performance at $\rho \leq 0.25$ is given by $R_c = 1.150$, the updating of the estimate and control limits is continued until the total number of sample inspected equals n^*, which is 100 000 for this case, provided that the process remains in control.

21.5 NUMERICAL EXAMPLES

In this section, examples based on the data in Table 21.6, taken from Table I in Xie *et al.*,[5] are presented. From the table, the first 20 data points are simulated from $p = 500$ ppm, after which the data is from $p = 50$ ppm.

Assuming that p_0 is given, the control limits of the chart can then be calculated directly, using $\phi = 0.006\,75$ and $\gamma_\phi = 1.306\,03$ (from Table 21.1). Figure 21.4 is the CCC chart plotted, given $p_0 = 500$ ppm. From the chart, it is observed that the chart signals at the 23rd observation, indicating the decrease in the process fraction nonconforming, p.

On the other hand, when p_0 is not given, by using the proposed scheme with sequential estimation, the estimation starts after m reaches 2 and the control limits can be obtained accordingly, with the values of ϕ_m and γ_{ϕ_m} given in Table 21.2. With ARL_0 specified at 200 (i.e. $\alpha_0 = 0.005$), $\rho \geq 2.5$, and $R = 1.1250$, m^* is 23 (from Table 21.3).

Table 21.7 shows the estimated p together with the control limits for sequential estimation. The CCC scheme is depicted in Figure 21.5, where the two dashed lines are the decision lines. From the chart, estimation is suspended at the third point as it is in the warning zone (beyond LDL). However, since there is no significant difference between the new sequential estimate from the subsequent data with $m = 3$ (104 ppm)

Table 21.6 A set of data for a simulated process (from Table I, Xie *et al.*[5]).

Nonconforming no.	CCC	Simulation legend	Nonconforming no.	CCC	Simulation legend
1	3 706	p = 500 ppm	16	753	p = 500 ppm
2	9 179	p = 500 ppm	17	3 345	p = 500 ppm
3	78	p = 500 ppm	18	217	p = 500 ppm
4	1 442	p = 500 ppm	19	3 008	p = 500 ppm
5	409	p = 500 ppm	20	3 270	p = 500 ppm
6	3 812	p = 500 ppm	21	5 074	Shift, p = 50 ppm
7	7 302	p = 500 ppm	22	3 910	p = 50 ppm
8	726	p = 500 ppm	23	23 310	p = 50 ppm
9	2 971	p = 500 ppm	24	11 690	p = 50 ppm
10	42	p = 500 ppm	25	19 807	p = 50 ppm
11	3 134	p = 500 ppm	26	14 703	p = 50 ppm
12	1 583	p = 500 ppm	27	4 084	p = 50 ppm
13	3 917	p = 500 ppm	28	826	p = 50 ppm
14	3 496	p = 500 ppm	29	9 484	p = 50 ppm
15	2 424	p = 500 ppm	30	66 782	p = 50 ppm

and the estimate before suspension (154 ppm), all data are combined. Estimation is suspended again at $M = 10$ and for similar reason; the estimation is resumed at $M = 20$. At $M = 22$, as there are four consecutive points plotted above the center line, estimation is suspended again.

The 23rd observation is plotted above the UCL, indicating possible process improvement. Thus, by using the proposed sequential estimation scheme with the guidelines given, the CCC chart with estimated parameter is as effective as that constructed assuming known p in detecting a change in p.

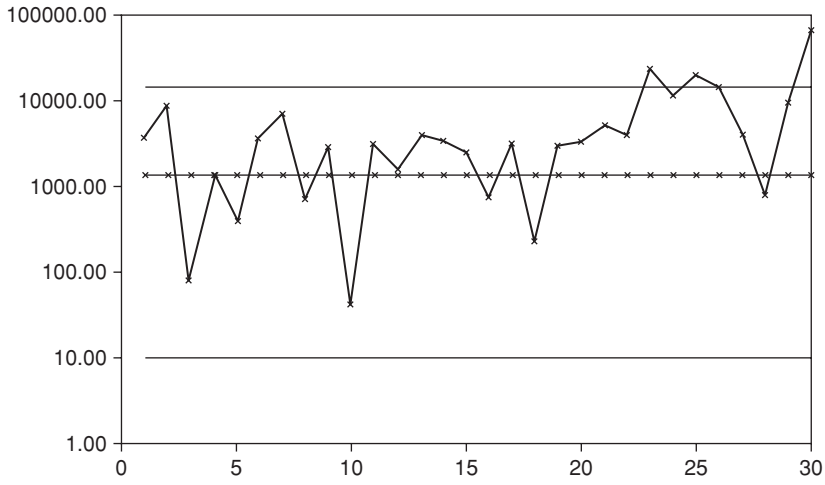


Figure 21.4 CCC chart when p_0 is known (= 500 ppm) and in-control ARL = 200.

Table 21.7 The values of $\bar{p}$ and the control limits with sequential estimator from Table 21.6.

No.	CCC	$\bar{p}$ (ppm)	LCL	UCL	No.	CCC	$\bar{p}$ (ppm)	LCL	UCL
1	3 706	–	–	–	16	753	303	12	23 023
2	9 179	78	31	105 092	17	3 345	303	12	23 189
3	78	**154**	19	51 689	18	217	303	12	21 923
4	1 442	154	19	51 689	19	3 008	303	12	21 988
5	4 09	154	19	51 689	20	3 270	347	11	21 864
6	3 812	268	12	29 410	21	*5 074*	334	12	22 693
7	7 302	231	15	33 847	22	*3 910*	**329**	13	23 024
8	726	263	13	29 651	23	*23 310*	329	13	23 024
9	2 971	270	13	28 709	24	*11 690*	329	13	23 024
10	42	**303**	12	25 462	25	*19 807*	329	13	23 024
11	3 134	303	12	25 462	26	*14 703*	329	13	23 024
12	1 583	303	12	25 462	27	*4,084*	329	13	23 024
13	3 917	303	12	25 462	28	*826*	329	13	23 024
14	3 496	303	12	25 462	29	*9 484*	329	13	23 024
15	2 424	303	12	25 462	30	*66 782*	329	13	22 897

If conventional estimation is used, $\hat{p}$ is 300 ppm, for an initial sample size $n = 20\,000$. With ARL_0 specified at 200, $\rho \geq 2.5$, and $R_c = 1.170$, $n^* = 50\,000$ (obtained by using equations (21.15) and (21.22)). Table 21.8 gives the values of the estimates together with the control limits. The estimate is updated once when the sample size collected reaches 50 000, which is the 18th observation in Table 21.8, as there is no out-of-control signal observed between the 7th and 18th plotted points.

This CCC scheme is depicted in Figure 21.6. The 23rd observation is plotted above the UCL, indicating a possible process improvement. Thus, by using this proposed scheme, the CCC chart is also able to detect a change in p.

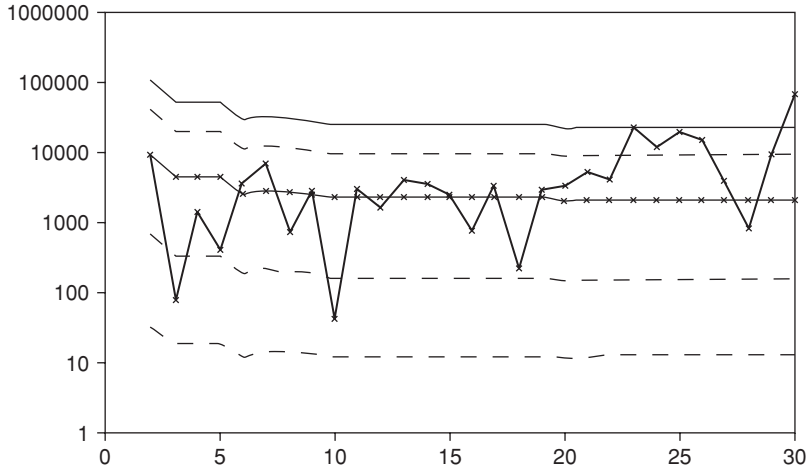


Figure 21.5 CCC chart under sequential estimation scheme simulated from initial $p_0 = 500$ ppm and $ARL_0 = 200$.

Table 21.8 The values of $\hat{p}$ and the control limits with conventionel estimator from Table 21.6.

No.	CCC	$\hat{p}$ (ppm)	LCL	UCL	No.	CCC	$\hat{p}$ (ppm)	LCL	UCL
1	3706	–	12	25965	16	753	300	12	25965
2	9179	–	12	25965	17	3345	300	12	25965
3	78	–	12	25965	18	217	**360**	12	21081
4	1442	–	12	25965	19	3008	360	12	21081
5	409	–	12	25965	20	3270	360	12	21081
6	3812	**300**	12	25965	21	*5074*	360	12	21081
7	7302	300	12	25965	22	*3910*	360	12	21081
8	726	300	12	25965	23	*23310*	360	12	21081
9	2971	300	12	25965	24	*11690*	360	12	21081
10	42	300	12	25965	25	*19807*	360	12	21081
11	3134	300	12	25965	26	*14703*	360	12	21081
12	1583	300	12	25965	27	*4084*	360	12	21081
13	3917	300	12	25965	28	*826*	360	12	21081
14	3496	300	12	25965	29	*9484*	360	12	21081
15	2424	300	12	25965	30	*66782*	360	12	21081

21.6 CONCLUSION

In this chapter, the basic properties of CCC charts as well as CCC schemes with estimated parameters are revisited. A set of comprehensive guidelines is given for the construction of CCC charts, when p_0 is known and when p_0 is estimated by two different schemes. In addition, the associated parameters for constructing CCC charts with the most commonly used in-control ARLs are given in Tables 21.1 and 21.2. Termination and suspension rules are introduced for the CCC scheme with sequentially

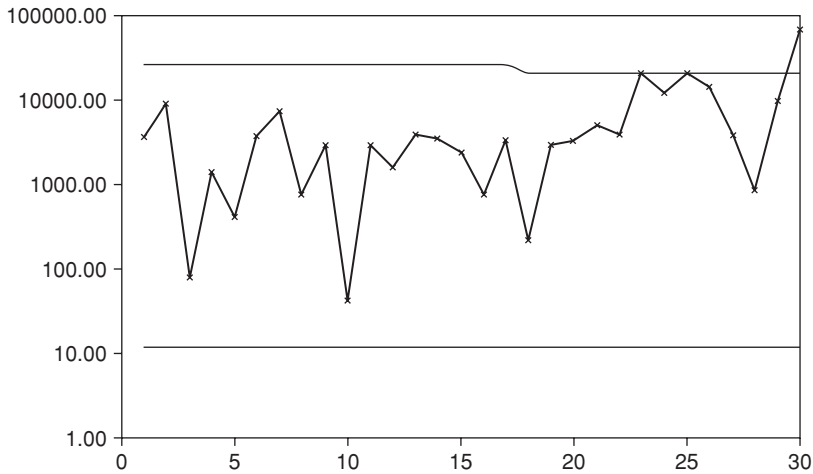


Figure 21.6 CCC chart with conventional estimation scheme simulated from $p_0 = 500$ ppm, and in-conrtol $ARL = 200$.

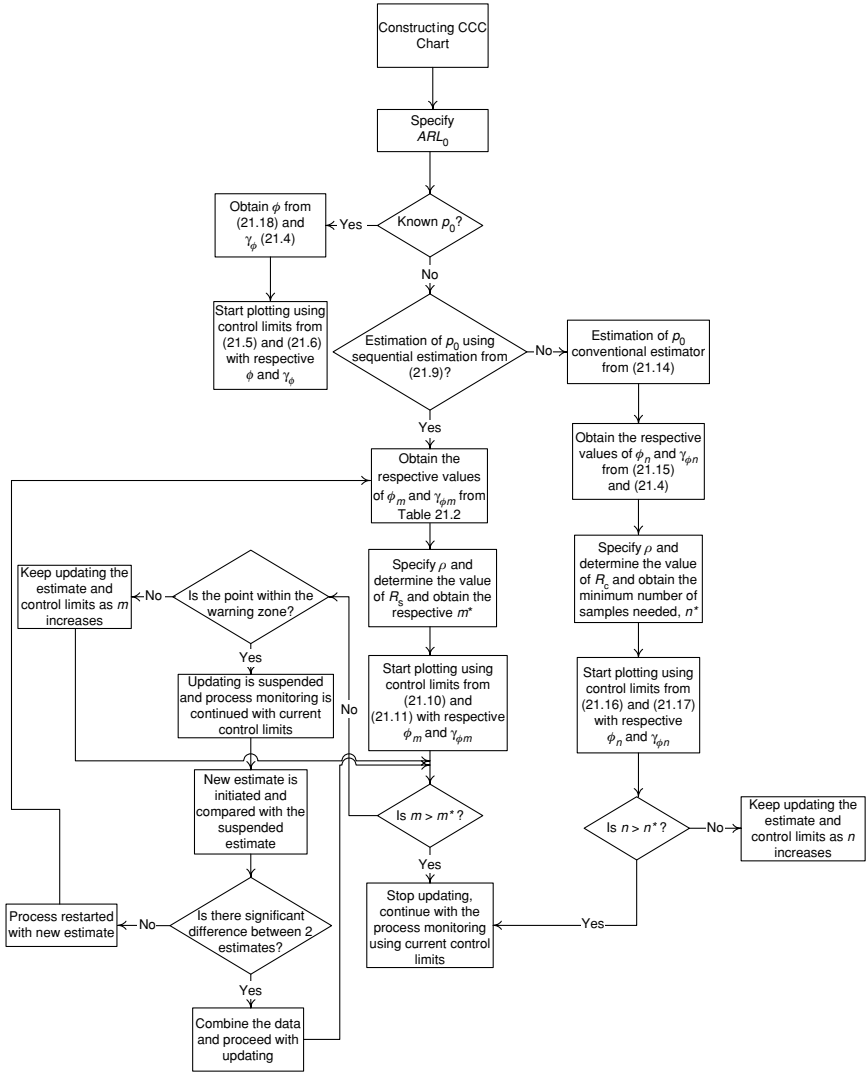

Constructing CCC Chart
Specify ARL_0
Known p_0?
Yes
Obtain ϕ from (21.18) and γ_ϕ (21.4)
No
Start plotting using control limits from (21.5) and (21.6) with respective ϕ and γ_ϕ
Estimation of p_0 using sequential estimation from (21.9)?
No
Estimation of p_0 conventional estimator from (21.14)
Yes
Obtain the respective values of ϕ_m and $\gamma_{\phi m}$ from Table 21.2
Obtain the respective values of ϕ_n and $\gamma_{\phi n}$ from (21.15) and (21.4)
Specify ρ and determine the value of R_s and obtain the respective m^*
Specify ρ and determine the value of R_c and obtain the minimum number of samples needed, n^*
Keep updating the estimate and control limits as m increases
No
Is the point within the warning zone?
Yes
Updating is suspended and process monitoring is continued with current control limits
Start plotting using control limits from (21.10) and (21.11) with respective ϕ_m and $\gamma_{\phi m}$
Start plotting using control limits from (21.16) and (21.17) with respective ϕ_n and $\gamma_{\phi n}$
No
New estimate is initiated and compared with the suspended estimate
Is $m > m^*$?
Is $n > n^*$?
No
Keep updating the estimate and control limits as n increases
Yes
Stop updating, continue with the process monitoring using current control limits
Yes
Process restarted with new estimate
No
Is there significant difference between 2 estimates?
Yes
Combine the data and proceed with updating

Figure 21.7 Flow chart for implementing CCC chart.

estimated parameter to enhance the sensitivity of the CCC scheme. An example is presented to illustrate the proposed scheme for constructing CCC charts. By way of summary, Figure 21.7 is a flowchart for constructing CCC charts for high-yield process monitoring when p_0 is known and when p_0 is estimated.

REFERENCES

1. Goh, T.N., and Xie, M. (2003) Statistical control of a Six Sigma process. *Quality Engineering*, 15, 587–592.
2. Glushkovsky, E.A. (1994) 'On-line' G-control chart for attribute data. *Quality and Reliability Engineering International*, 10, 217–227.

3. Xie, W., Xie, M., and Goh, T.N. (1995) A Shewhart-like charting technique for high yield processes. *Quality and Reliability Engineering International*, 11, 189–196.
4. Kuralmani, V., Xie, M., and Goh, T.N. (2002) A conditional decision procedure for high yield processes. *IIE Transactions*, 34, 1021–1030.
5. Tang, L.C. and Cheong, W.T. (2004) Cumulative conformance count chart with sequentially estimated parameter. *IIE Transactions*, 36, 841–853.
6. Yang, Z., Xie, M., Kuralmani, V., and Tsui, K.L. (2002) On the performance of geometric charts with estimated control limits. *Journal of Quality Technology*, 34, 448–458.
7. Xie, M., Goh, T.N., and Kuralmani, V. (2000) On optimal setting of control limits for geometric chart. *International Journal of Reliability, Quality and Safety Engineering*, 7, 17–25.
8. Zhang, L., Govindaraju, K., Bebbington, M., and Lai, C.D. (2004) On the statistical design of geometric control charts. *Quality Technology and Quantitative Management*, 1, 233–243.
9. Haldane, J.B.S. (1945) A labour-saving method of sampling. *Nature*, 155, 3924.

22

Simultaneous Monitoring of the Mean, Variance and Autocorrelation Structure of Serially Correlated Processes

O. O. Atienza and L. C. Tang

Statistical process control techniques for monitoring serially correlated or autocorrelated processes have received significant attention in the statistical quality engineering literature. The focus of most studies, however, is on the detection of the mean shift of the process. The detection of changes in the variance and autocorrelation structure of the series due to some special causes of variation affecting the system is often overlooked. Ideally, one needs to maintain three control charts in order to effectively detect changes in the process: one each for detecting change in the mean, variance and autocorrelation structure (MVAS) of a series. Such an approach can be quite cumbersome to implement. One alternative is to develop a monitoring scheme that has good sensitivity in simultaneously detecting changes in the MVAS of a series. An example of such an approach is the sample autocorrelation chart (SACC) proposed by Montgomery and Friedman[1]. Unfortunately, it has been found that the SACC has poor sensitivity in detecting shifts in the mean and variance. As an alternative, we propose a new monitoring scheme based on the characteristics of a Gaussian stationary time series. Compared to the SACC, the proposed scheme is simpler to implement and is more sensitive to changes in the MVAS of a series.

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

22.1 INTRODUCTION

Statistical process control (SPC) techniques aim to detect as early as possible the presence of external or special causes of variation affecting a given process. These special causes of variation usually result in unwanted deviation of important product characteristics from the desired or target value. SPC procedures rely mainly on the assumption that, when the process is affected only by the inherent or common causes of variation, the process will produce measurements that are independent and identically distributed over time. Any deviation from such in-control iid behavior is usually interpreted as an indication of the presence of special causes of variation. This assumption has been the mainstay of classical SPC techniques.

Recent advances in manufacturing technology, however, are challenging the applicability of the in-control iid concept.[1] For example, in continuous chemical processing and automated manufacturing environments, process measurements are often serially correlated even when there are no special causes of variation affecting the system. This autocorrelation violates the in-control independence assumption in many SPC control charting procedures. Such a violation has a significant impact on the performance of the classical SPC procedures.[2,3]

The most popular approach for monitoring processes with serially correlated observations is to model the inherent autocorrelation of the process measurements using an autoregressive moving average (ARMA) model. To detect changes in the process, residuals are generated using the chosen model. When the model is appropriately chosen and well estimated, these residuals approximate an iid behavior. Using this assumption, we can employ the traditional SPC charts to monitor the residuals. Any deviation of the residual from iid behavior indicates a change in the process that must be due to a special cause of variation.

The development of SPC techniques for monitoring autocorrelated processes has received considerable attention in the quality engineering literature. The focus, however, is mainly on the detection of the mean shift of the process. The detection of changes in the variance and autocorrelation structure of the series, which are also important indicators of the presence of process changes that must be due to a special cause of variation affecting the system, is often overlooked. There are, however, some notable exceptions.[4–9]

Dooley and Kapoor[4] suggested monitoring of the changes in the mean, variance and autocorrelation structure (MVAS) of the process measurements by simultaneously maintaining a CUSUM, a χ^2, and an autocorrelation chart on residuals. Yourstone and Montgomery[5] noted that the autocorrelation function (ACF) 'will detect shifts in the autocorrelation structure as well as shifts in the mean and variance of the real-time process quality data'. They suggested real-time monitoring of the first m residual autocorrelations calculated for the n latest process observations. They call their chart the sample autocorrelation chart or (SACC). Atienza *et al.*[8] studied the average run length (ARL) performance of the SACC. They noted that although the SACC can detect changes in the mean and variance of a series, it does not perform well even in comparison with the Shewhart control chart (SCC) on residuals. The only advantage in using the SACC is that it can detect changes in the autocorrelation structure of a series better than the SCC. Thus, one cannot rely on the SACC to simultaneously monitor changes in the MVAS of a series. In this chapter, we propose an alternative to the SACC. Compared to the SACC, the proposed procedure, which is based on the

characteristics of a stationary Gaussian process, is much simpler to implement and is sensitive to shifts in the MVAS of a series. Although the proposed procedure can be implemented for the general ARMA case, we will focus our analysis on the important case of AR(1) process. The importance of the AR(1) process in process control has been emphasized in the literature.[2,10–12]

22.2 THE PROPOSED APPROACH

We follow the usual assumption that when the process is stable, the sequence of process measurements $\{x_t\}$ can be modeled as a stationary Gaussian ARMA(p, q) process

$$\phi_p(B)x_t = \phi_0 + \theta_q(B)\varepsilon_t,$$

where $\{x_t\}$ is a stationary time series representing the process measurements,

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

is an autoregressive polynomial of order p,

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

is a moving average polynomial of order q, B is the backshift operator, and $\{\varepsilon_t\}$ is a sequence of normally and independently distributed random errors with mean zero and constant variance σ_ε^2. Without loss of generality, we assume that the level ϕ_0 of the in-control process measurements is zero.

By definition, if $\{x_t, t = 1, 2, 3, \ldots\}$ is a stationary Gaussian process then $\{x_t\}$ is strictly stationary, since for all $n \in \{1, 2, \ldots\}$ and for all $h, t_1, t_2, \ldots \in Z$, the random vectors $(x_{t_1}, \ldots, x_{t_n})'$ and $(x_{t_1+h}, \ldots, x_{t_n+h})'$ have the same mean and covariance matrix, and hence the same distribution (see p. 13 of Brockwell and Davis[13]). Let $\{x_t\}$ be a stationary Gaussian ARMA process. Then $\mathbf{x} = (x_{t_1}, \ldots, x_{t_n})'$ is multivariate normal with mean μ_0 and covariance matrix

$$\Sigma_0 = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{bmatrix} = \sigma_x^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix},$$

where γ_k and ρ_k represent the autocovariance and autocorrelation at lag k, respectively. For an ARMA (1,1) process the autocovariance function is given by

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_\varepsilon^2,$$

$$\gamma_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma_\varepsilon^2,$$

$$\gamma_k = \phi_1\gamma_{k-1}, \qquad k \geq 2,$$

while the autocorrelation function is given by

$$\rho_k = \begin{cases} 1, & k = 1, \\ \phi_1 \rho_{k-1}, & k \geq 2. \end{cases}$$

The autocovariance and autocorrelation function of the general ARMA(p, q) process can be found on p. 76 of Box *et al.*[14]. The procedures for estimating the autocovariance function are discussed in detail by Brockwell and Davis[13] and Box et al.[14]

From the above characteristic of the Gaussian ARMA process, we know that

$$H_n = \mathbf{x}' \Sigma_0^{-1} \mathbf{x} \tag{22.1}$$

has a χ^2 distribution with n degrees of freedom. This statistic is sensitive to changes in both the mean vector and the autocovariance matrix. If we divide $\{x_1, x_2, x_3, \ldots\}$ into disjoint series of measurements of length n ($\{\mathbf{x}'_1 = (x_1, x_2, \ldots, x_n), \mathbf{x}'_2 = (x_{n+1}, x_{n+1}, \ldots, x_{2n}), \ldots\}$), we can choose n large enough such that the correlation among $\{\mathbf{x}'_1, \mathbf{x}'_2, \ldots\}$ becomes negligible. Hence, we can implement the typical χ^2 control chart for monitoring an ARMA process. Since this control chart is analogous to the multivariate T^2 control chart for monitoring products or processes with several correlated quality characteristics,[10,15] one may extend the application to the multivariate CUSUM or exponentially weighted moving average (EWMA) for monitoring autocorrelated processes.[16–19]

In implementing the above χ^2 control chart, after calculating an H statistic, one needs to wait for all the following n observations before computing the next H statistic. This can slow down the detection of process changes, especially when n is large. One alternative is to calculate the H statistic for the latest n observations, that is, using the following sequence of vector of measurements:

$$\mathbf{x}_1^* = (x_1, x_2, \ldots, x_n)', \mathbf{x}_2^* = (x_2, x_3, \ldots, x_{n+1})', \mathbf{x}_3^* = (x_3, x_4, \ldots, x_{n+2})', \ldots.$$

In what follows, we will call the H chart based on the above sequence of measurements the moving H chart or MH chart. For this case, the resulting H statistics are no longer independent. Thus we cannot establish the control limits using an approach similar to the construction of the T^2 control chart. We can use simulation to determine the control limits that will achieve a desired in-control ARL.

22.3 ARL PERFORMANCE

The monitoring strategy described in the previous section is obviously a Shewhart scheme. When $n = 1$, one can easily verify that the proposed control charting strategy becomes a special case of a Shewhart control chart with modified control limits. Thus, in the ensuing sections we will only compare the ARL performance of the proposed scheme with similar Shewhart monitoring schemes (SCC and its rival SACC).

22.3.1 Sensitivity in detecting a mean shift

The sensitivity of the MH chart is dependent on the size of the moving window n. Using Monte Carlo simulation, we analyze the effect of n on the ability of the MH

chart to detect a change in the mean of an AR(1) process. The shift in the mean is expressed in terms of σ_x,

$$\sigma_x = \sqrt{\frac{\sigma_\varepsilon^2}{1-\phi_1^2}},$$

where ϕ_1 represents the autoregressive parameter. Each simulation run is composed of 5000 iterations. The required ARMA process is generated using the innovation algorithm,[13] while the sequence of iid normal random variables is generated using the IMSL[20] Statistical Library. For comparison purposes, all the control limits of the *MH* chart in the following are chosen such that the in-control ARL is approximately 370. The effect of n on the sensitivity of the *MH* chart in detecting shifts in the mean is shown in Figure 22.1. It will be evident that when the autocorrelation is low we need large n to detect small changes in the mean. However, when the process is highly positively autocorrelated an n close to 5 gives good sensitivity in detecting both small and large shifts in the mean.

Figure 22.2 compares the ARL performance of the SCC, SACC, and *MH* charts in detecting a shift in the mean of an AR(1) process. The ARL figures for the SCC chart are obtained using the program published by Wardell *et al.*[21] The SACC is based on the first 15 autocorrelations of the latest 200 residuals. The control limit constant D is set at 2.835 and $\lambda = 1$. The ARL figures for the SACC are calculated using Monte Carlo simulation. The resulting ARLs are shown in Figure 22.2. Except for mean shifts that can produce large residual values, it will be evident that the *MH* chart is more

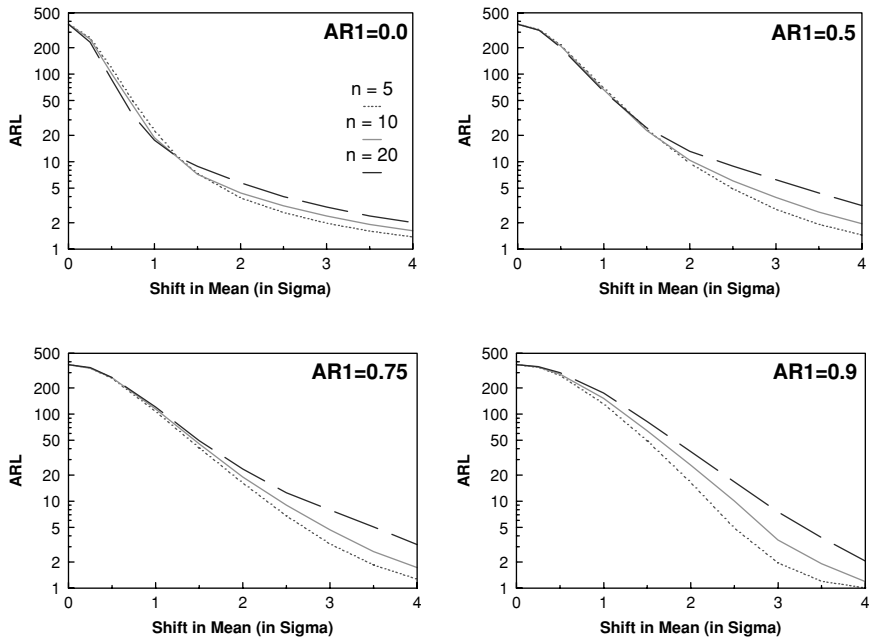


Figure 22.1 Effect of changing n on the sensitivity of the *MH* chart in deleting mean shift.

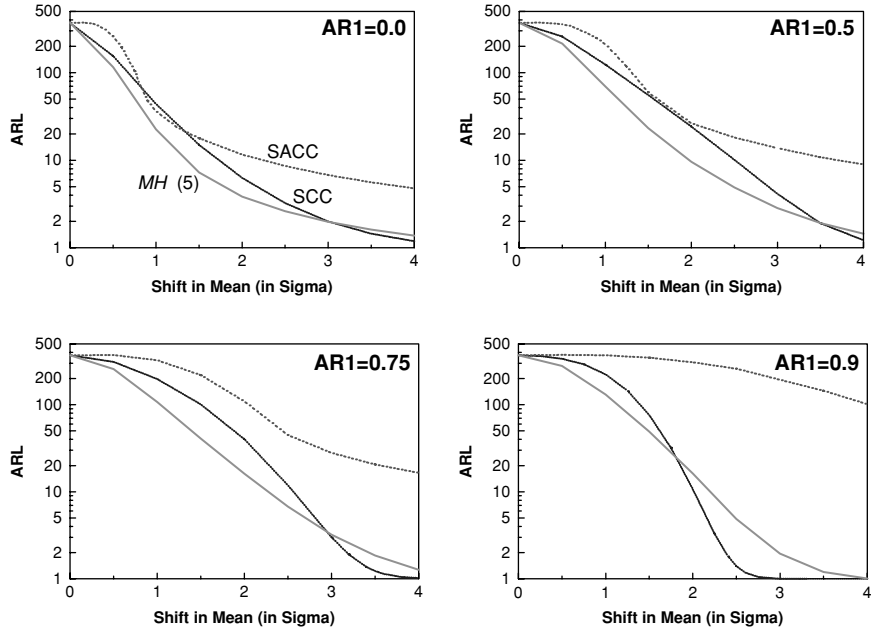


Figure 22.2 ARL performance of the SCC, SACC and *MH* chart in delecting shifts in the process mean.

sensitive in detecting a mean shift than the SCC. In all cases, the *MH* chart outperforms its main rival, the SACC.

22.3.2 Sensitivity in detecting a variance shift

Although the SACC can react to changes in the process variance, it is very insensitive.[8] Analogous to the Shewhart control charts for individuals, which contain both the information concerning the mean and the variance of the series, the SCC has better ARL performance in detecting a shift in the variance than the SACC. Extending the comparison to include the proposed chart, we can see from Figure 22.3 that the *MH* chart achieves the best ARL performance in monitoring changes in the variance.

22.3.3 Sensitivity in detecting changes in the autocorrelation structure

Another important indicator of the presence of an external cause of variation affecting the process is the change in the autocorrelation structure of $\{x_t\}$. Assuming that the process starts in an in-control iid $N(0,1)$ mode, Figure 22.4 compares the sensitivity of the SCC, SACC and *MH* chart in detecting changes in the autocorrelation structure of a series. The SACC is evidently superior in detecting small shifts in the AR(1) or MA(1) parameter. For large shifts, however, the *MH* chart is more sensitive. These results are expected since the SACC is based on a large number of samples while the

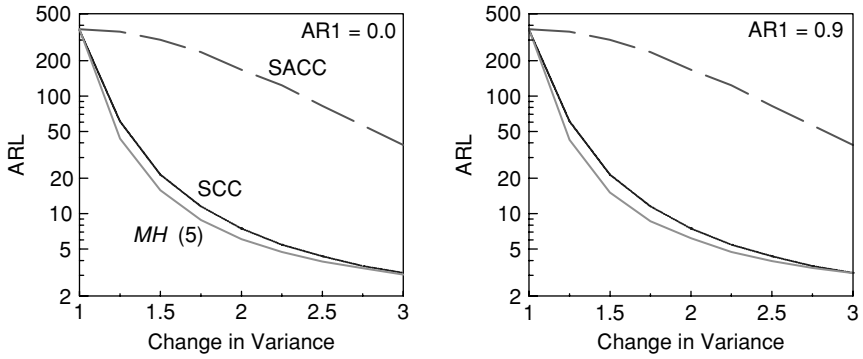


Figure 22.3 Sensitivity of the SCC, SACC and *MH* chart in delecting shifts in the variance.

MH chart analyzed in Figure 22.4 is just based on $n = 5$. The *MH* chart is also superior to the SCC.

22.4 NUMERICAL EXAMPLE

We demonstrate the application of the proposed control charting technique by simulating changes in an AR(1) series. To represent an in-control process state, we generated 200 observations from an AR(1) process with $\phi_1 = 0.9$, $\mu = 0$ and $\sigma_\varepsilon = 1$. Starting at $t = 201$, we introduced a mean shift of size $1.5\sigma_x$. The mean moved back to zero at $t = 301$ but the variance increased by 20%. At $t = 401$, the error variance returned to the original level but the process became an ARMA(1,1) with $\phi_1 = 0.9$ and $\theta = 0.6$. The resulting series is shown in Figure 22.5.

To illustrate the mechanics of calculating the statistic plotted in an *MH* chart, the process observations for $t = 196$ to $t = 205$ are shown in Table 22.1. With $n = 5$, the control limit that will produce an in-control ARL of 370 is approximately 16.32.

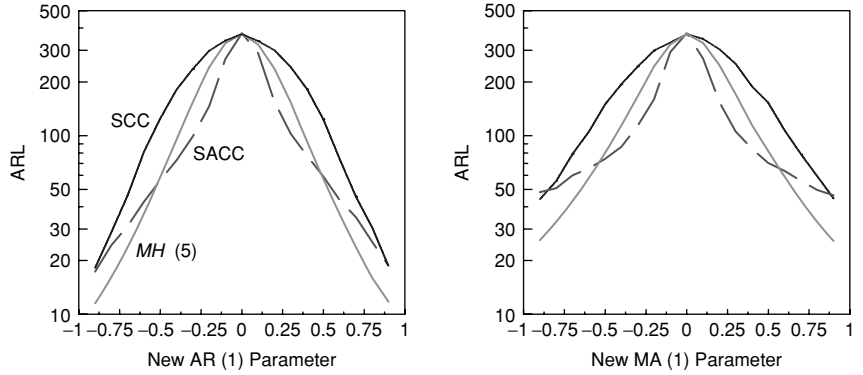


Figure 22.4 Sensitivity of the SCC, SACC and the *MH* chart in delecting shifts in model parameters.

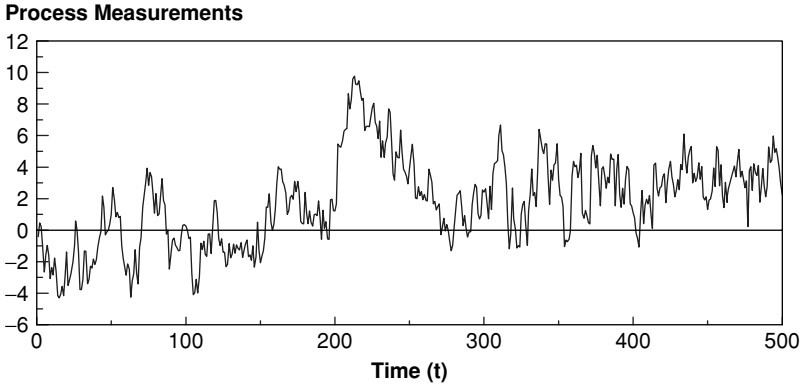


Figure 22.5 The simulated process measurements.

The *MH* statistic at $t = 200$ must be based on the vector of observations $\mathbf{x}^*_{200} = (x_{196}, x_{197}, \ldots, x_{200})'$. For the in-control state, the corresponding covariance matrix and its inverse are as follows:

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 5.263 & 4.737 & 4.263 & 3.837 & 3.453 \\ 4.737 & 5.263 & 4.737 & 4.263 & 3.837 \\ 4.263 & 4.737 & 5.263 & 4.737 & 4.263 \\ 3.837 & 4.263 & 4.737 & 5.263 & 4.737 \\ 3.453 & 3.837 & 4.263 & 4.737 & 5.263 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_0^{-1} = \begin{bmatrix} 1.00 & -0.90 & 0.00 & 0.00 & 0.00 \\ -0.90 & 1.81 & -0.90 & 0.00 & 0.00 \\ 0.00 & -0.90 & 1.81 & -0.90 & 0.00 \\ 0.00 & 0.00 & -0.90 & 1.81 & -0.90 \\ 0.00 & 0.00 & 0.00 & -0.90 & 1.81 \end{bmatrix}.$$

Table 22.1 Simulated process measurements, $196 \leq t \leq 205$.

t	x_t	*MH* statistic	t	x_t	*MH* statistic
196	−0.577	2.644*	201	1.666	1.383
197	1.934	7.311*	202	5.470	17.128
198	1.968	7.226*	203	5.315	16.546
199	1.234	6.426*	204	5.265	16.761
200	1.224	6.437	205	5.672	17.555

*Calculated using observations not shown in this table.

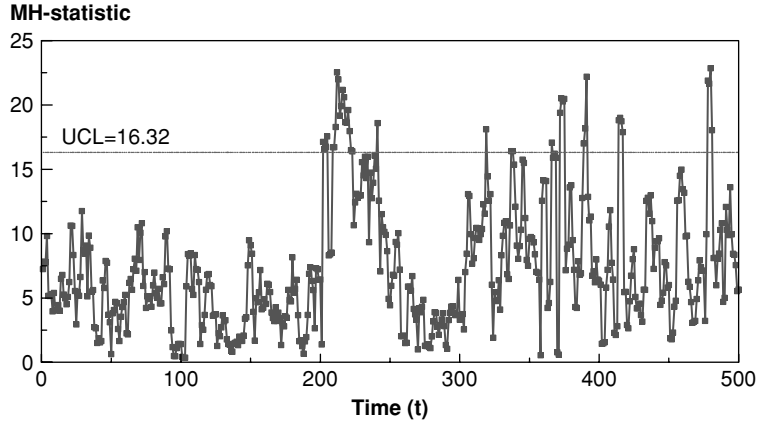


Figure 22.6 The *MH* chart for the series in Figure 22.5.

At $t = 200$, the *MH*-statistic is therefore given by

$$MH = [-0.577\ 1.934\ 1.968\ 1.234\ 1.224] \begin{bmatrix} 1.00 & -0.90 & 0.00 & 0.00 & 0.00 \\ -0.90 & 1.81 & -0.90 & 0.00 & 0.00 \\ 0.00 & -0.90 & 1.81 & -0.90 & 0.00 \\ 0.00 & 0.00 & -0.90 & 1.81 & -0.90 \\ 0.00 & 0.00 & 0.00 & -0.90 & 1.00 \end{bmatrix} \begin{bmatrix} -0.577 \\ 1.934 \\ 1.968 \\ 1.234 \\ 1.224 \end{bmatrix}$$

$$= 6.437$$

The above quantity be can easily obtained using standard spreadsheet software such as Excel or Lotus1-2-3. The next *MH*-statistic, which is based on the vector of observations $\mathbf{x}^*_{201} = (x_{197}, x_{198}, \ldots, x_{201})'$, can be computed in the same manner. The *MH* statistics for $t = 196$ to $t = 205$ are given in Table 22.1. The entire control chart for the series in Figure 22.5 is shown in Figure 22.6. Clearly, the proposed control chart effectively detected the simulated process excursions.

22.5 CONCLUSION

Most modern industrial processes produce measurements that are autocorrelated. To effectively detect the presence of process excursions caused by external sources of variation, we must simultaneously monitor changes in the MVAS of the process measurements. One way to monitor an autocorrelated process is to maintain three control charts (one each for detecting change in the mean, variance, and autocorrelation structure of the series). Another alternative is to devise a monitoring procedure that can simultaneously detect changes in the MVAS. Using the characteristics of a stationary Gaussian ARMA process, we develop a control-charting scheme based on the H statistic (22.1). In comparison with other Shewhart schemes for monitoring autocorrelated processes, the proposed moving H chart is found to be effective in simultaneously detecting shifts in the MVAS of a series.

REFERENCES

1. Montgomery, D.C, and Friedman, D.J, (1989) Statistical process control in computer-integrated manufacturing environment. In J.B. keats and N.F. Hubele (eds). *Statistical Process Control in Automated Manufacturing*. New york: Marcel Dekker.
2. Harris, T.J. and Ross, W.H. (1991) Statistical process control procedures for correlated observations. *Canadian Journal of Chemical Engineering*, 69, 48–57.
3. Alwan, L.C. (1992) Effects of autocorrelation on control chart performance. *Communications in Statistics: Theory and Methods*, 21, 1025–1049.
4. Dooley, K.J. and Kapoor, S.G. (1990) An enhanced quality evaluation system for continuous manufacturing processes, Part 1: Theory. *Transactions of the ASME, Journal of Engineering for Industry*, 112, 57–62.
5. Yourstone, S.A. and Montgomery, D.C. (1989) A time-series approach to discrete real-time process quality control. *Quality and Reliability Engineering International*, 5, 309–317.
6. Yourstone, S.A. and Montgomery, D.C. (1991) Detection of process upsets – sample autocorrelation control chart and group autocorrelation control chart applications. *Quality and Reliability Engineering International*, 7, 133–140.
7. Lu, C.W. and Reynolds, M.R. (1999) Control charts for monitoring the mean and variance of autocorrelated processes. *Journal of Quality Technology*, 31, 259–274.
8. Atienza, O.O., Tang, L.C. and Ang, B.W. (1997) ARL properties of a sample autocorrelation chart. *Computers and Industrial Engineering*, 33, 733–736.
9. Atienza, O.O., Tang, L.C. and Ang, B.W. (2002) Simultaneous monitoring of sample and group autocorrelations. *Quality Engineering*, 14, 489–499.
10. Montgomery D.C. (1996) *Introduction to Statistical Quality Control*, 3rd edn. New York: John Wiley & Sons, Inc.
11. Wardell, D.G., Moskowitz, H. and Plante, R.D. (1994) Run-length distributions of special-cause control charts for correlated processes (with discussion). *Technometrics*, 36, 3–27.
12. Runger, G.C., Willemain, T.R. and Prabhu, S. (1995) Average run lengths for CUSUM control charts applied to residuals. *Communications in Statistics: Theory and Methods*, 24, 273–282.
13. Brockwell, P.J. and Davis, R.A. (1987) *Time Series Analysis: Theory and Methods*. New York: Springer-Verlag.
14. Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) *Time Series Analysis, Forecasting and Control*. Englewood Cliffs, NJ: Prentice Hall.
15. Alt, F.B. (1985) Multivariate quality control. in S. Kotz and N.L. Johnson (eds), *Encyclopedia of Statistical Sciences*, Vol. 6. New York: John Wiley & Sons, Inc.
16. Healy, J.D. (1987) A note on multivariate CUSUM procedures. *Technometrics*, 29, 409–412.
17. Crosier, R.B. (1988) Multivariate generalization of cumulative sum quality-control schemes. *Technometrics*, 30, 291–303.
18. Lowry, C.A., Woodall, W.H., Champ, C.W. and Rigdon, S.E. (1992) Multivariate exponentially weighted moving average control chart. *Technometrics*, 34, 46–53.
19. Prabhu, S.S. and Runger, G.C. (1997) Designing a multivariate EWMA control chart. *Journal of Quality Technology*, 29, 8–15.
20. IMSL (1991) *User's Manual Statistical Library*, Version 2. Houston, TX: IMSL.
21. Wardell, D.G., Moskowitz, H. and Plante, R.D. (1994) Run-length distributions of residual control charts for autocorrelated processes. *Journal of Quality Technology*, 26, 308–317.
22. Roes, K.C.B., Does, R.J.M.M. and Schurink, Y. (1993) Shewhart-type control charts for individual observations. *Journal of Quality Technology*, 25, 188–198.

23

Statistical Process Control for Autocorrelated Processes: A Survey and An Innovative Approach

L. C. Tang and O. O. Atienza

In this chapter, we first give a quick survey of the work on statistical process control for autocorrelated processes, which is still under very intensive research. We then present an approach involving monitoring some statistics originating from time series models. The performance of the proposed monitoring scheme is evaluated and compared with the conventional method based on monitoring the residuals of a time series model.

23.1 INTRODUCTION

Traditional statistical process control (SPC) schemes, such as Shewhart and cumulative sum (CUSUM) control charts, assume that data collected from the process are independent. However, this assumption has been challenged as it has been found that, in many practical situations, data are serially correlated. The performance of traditional control charts deteriorates significantly under autocorrelation. This motivated the pioneering work by Alwan and Roberts,[1] who proposed the monitoring of forecasted errors after an appropriate time-series model has been fitted to the process. This method is intuitive as autocorrelation can be accounted for by the underlying time series model while the residual terms capture the independent random errors of the process. Traditional SPC schemes can be applied to monitoring the residuals. Subsequent work on this problem can be broadly classified into two themes; those

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

based on time series models[1–8] and those which are model-free.[6,19–28] For the former, three general approaches have been proposed: those which monitor residuals,[1–18,16,17] those based on direct observations,[9–14,18] and those based on new statistics.[15] A brief account of these approaches is presented in this chapter.

Wardell *et al.*[2] and Lu and Reynolds[3,4] proposed the use of exponentially weighted moving average (EWMA) control charts for monitoring residuals. Apley and Shi[5] proposed a generalized likelihood ratio test (GLRT) approach in detecting the mean shift in autocorrelation processes. Apley and Tsung[6] proposed a triggered cumulative score (cuscore) chart, which has similar character to GLRT but is easier to implement. Castagliola and Tsung[7] studied the impact of non-normality on residual control charts. They proposed a modified Shewhart control chart for residuals – the special cause chart (SCC) – which is more robust to non-normal situations. Testik[8] considered the uncertainty in the time series model due to estimated parameters and proposed wider EWMA control limits for monitoring the residuals of the first-order autoregressive AR(1) process.

Taking a different approach involving direct monitoring of process outputs instead of the residuals, Montgomery and Mastrangelo[9] proposed the moving centerline EWMA (MCEWMA) chart, which approximates the time series model with the EWMA model. Mastrangelo and Brown[10] provided a further study of the properties of the MCEWMA chart. Alwan[11] developed a general strategy to study the effects of autocorrelation using the Shewhart control chart. Timmer *et al.*[12] compared different CUSUM control charts for AR(1) processes. Lu and Reynolds[13] study the performance of CUSUM control charts for both residuals and observations, and concluded that they are almost equally effective. The authors claimed that the performance of CUSUM and EWMA charts in detecting mean shift is comparable. Atienza, *et al.*[14] also proposed a CUSUM scheme for autocorrelated observations. They also[15] proposed a new test statistic for detecting the additive outlier (AO), innovational outlier (IO), and level shift (LS).

The time series model based approach is easy to understand and effective in some situations. However, it requires identification of an appropriate time series model from a set of initial in-control data. In practice, it may not be easy to establish and may appear to be too complicated to practicing engineers. Hence, the model-free approach has recently attracted much attention. The most popular model-free approach is to form a multivariate statistic from the autocorrelated univariate process, and then monitor it with the corresponding multivariate control chart.

Krieger *et al.*[19] used a multivariate CUSUM scheme. Apley and Tsung[6] adapted the T^2 control chart for monitoring univariate autocorrelated process. Atienza *et al.*[20] proposed a Multivariate boxplot-T^2 control chart. Dyer *et al.*[21] adapted the use of the multivariate EWMA control chart for autocorrelated processes.

Another model-free approach is to use a batch means control chart, proposed by Runger and Willemain,[22] referenced by Montgomery,[29] and discussed in detail by Sun and Xu.[23] The main advantage of this approach lies in its simplicity. In an attempt to 'break' dependency, it simply divides sequential observations into a number of batches, and then monitors the means of these batches on a standard individuals control chart.

Other works on SPC of autocorrelated processes includes Balkin and Lin[24] who studied the use of sensitizing rules for the Shewhart control chart on autocorrelated

data; and Zhang,[25] Noorossana and Vaghefi[30] and Kalgonda and Kulkarni[31] who looked into control charts for multivariate autoregressive processes.

Although much work has been done on statistical control of autocorrelated processes, there has hitherto been no single method that is widely accepted or proven to be effective in most situations. It is our opinion that the most promising approach is the use of a newly conceived statistic. In the following, we present the scheme proposed by Atienza *et al.*[15] for monitoring autocorrelated processes. We focus on studying the behavior of the statistic $\lambda_{\mathrm{LS,t}}$ used for detecting level shifts in an AR(1) process. The performance of the scheme is evaluated and a comparison with Alwan and Roberts's SCC is made.

23.2 DETECTING OUTLIERS AND LEVEL SHIFTS

Consider the ARMA model

$$\phi(B)Z_t = \phi_0 + \theta(B)\varepsilon_t, \tag{23.1}$$

where Z_t is a stationary time series representing the process measurements,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

is an autoregressive polynomial of order p,

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

is a moving average polynomial of order q, B is the backshift operator, and $\{\varepsilon_t\}$ is a sequence of normally and independently distributed random errors with mean zero and constant variance σ^2. Without loss of generality, we assume that the level ϕ_0 of the time series $\{Z_t\}$ is zero. If we let $\hat{Z}_t$ represent the predicted value obtained from an appropriately identified and fitted ARMA model, then the residuals $\{e_1 = Z_1 - \hat{Z}_1, e_2 = Z_2 - \hat{Z}_2, \ldots, e_t = Z_t - \hat{Z}_t, \ldots\}$ will behave like independent and identically distributed (i.i.d.) random variables.[32]

Let

$$Y_t = f(t) + \frac{\theta(B)}{\phi(B)}\varepsilon_t, \tag{23.2}$$

where Y_t and $f(t)$ represent the 'contaminated' series and the anomalous exogenous disturbances such as outliers and level shifts, respectively. The function $f(t)$ may be deterministic or stochastic, depending on the type of disturbance. For a deterministic model, $f(t)$ is of the form

$$f(t) = \omega_0 \frac{\omega(B)}{\delta(B)} \xi_t^{(d)}, \tag{23.3}$$

where

$$\xi_t^{(d)} = \begin{cases} 1, & \text{if } t = d, \\ 0, & \text{if } t \neq d, \end{cases}$$

is an indicator variable for the occurrence of a disturbance at time d, $\omega(B)$ and $\delta(B)$ are backshift polynomials describing the dynamic effect of the disturbance on Y_t, and

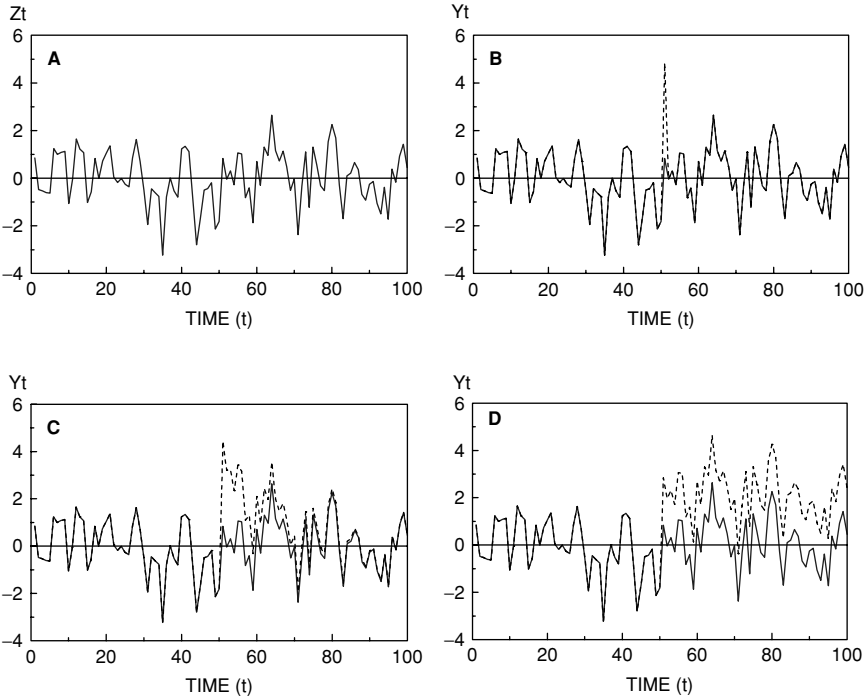


Figure 23.1 (A) A simulated AR(1) series with $\phi = 0.5$. (B) The simulated AR(1) series with AO at $t = 51$. (C) The simulated AR(1) series with IO at $t = 51$. (D) The simulated AR(1) series with LS at $t = 51$.

ω_0 is a constant denoting the initial impact of the disturbance. When $\omega(B)/\delta(B) = 1$, the disturbance is an additive outlier. An AO affects the level of the observed time series only at time $t = d$. A common cause of AOs is data recording error. In a discrete manufacturing process, an AO can occur when there are mixed units in a large lot of raw materials. When $\omega(B)/\delta(B) = \theta(B)/\phi(B)$, equation (23.3) represents an innovational outlier. An IO affects the level of Y_t at $t = d$. After $t = d$, this effect fades exponentially. An IO is most likely caused by a contaminant in a continuous chemical process. During preventative maintenance in a chemical factory, for example, if an existing spare part (e.g. pipe or other connector) is replaced by a contaminated unit, the characteristics of the chemical being processed will be grossly affected at time $t = d$ but thereafter the effect of the contaminant will fade. When $\omega(B)/\delta(B) = 1/(1 - B)$, the disturbance represented by equation (23.3) is a level shift. An LS shifts up or down the level of Y_t starting at $t = d$. This shift persists for $t > d$. An LS is mainly caused by a change in material quality or process settings. Figure 23.1 illustrates the effect of the three types of outlier on an AR(1) process. In what follows, we use ω_{AO}, ω_{IO}, and ω_{LS} to distinguish whether ω_0 is associated with an AO, IO or LS, respectively.

From equations (23.2) and (23.3) we obtain the following expression:

$$\frac{\phi(B)}{\theta(B)} Y_t = \frac{\phi(B)\omega(B)}{\theta(B)\delta(B)} \xi_t^{(d)} \omega_0 + \varepsilon_t.$$

If we let $y_t = [\phi(B)/\theta(B)]Y_t$ and $x_t = [\phi(B)\omega(B)/\theta(B)\delta(B)]\xi_t^{(d)}$, we have

$$y_t = \omega_0 x_t + \varepsilon_t,$$

which is just a simple linear regression equation. Thus, the impact parameter ω_0 can be estimated using

$$\hat{\omega}_0 = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2} \qquad \text{with } \mathrm{Var}(\hat{\omega}_0) = \frac{\sigma^2}{\sum_{t=1}^{T} x_t^2},$$

where T represents the sample size. Using these equations, we can obtain the following estimates of ω_0 for the three types of disturbance mentioned above:

$$\hat{\omega}_{\mathrm{AO},t} = \begin{cases} \rho^2_{\mathrm{AO},t}\left(y_t - \sum_{\mathrm{i=1}}^{\mathrm{T-t}} \pi_i y_{t+i}\right) & t = 1, 2, \ldots, T-1, \\ y_t & t = T, \end{cases} \tag{23.4}$$

$$\hat{\omega}_{\mathrm{IO},t} = y_t \qquad t = 1, 2, \ldots, T, \tag{23.5}$$

$$\hat{\omega}_{\mathrm{LS},t} = \begin{cases} \rho^2_{\mathrm{LS},t}\left(y_t - \sum_{\mathrm{i=1}}^{\mathrm{T-t}} \eta_i y_{t+i}\right) & t = 1, 2, \ldots, T-1, \\ y_t & t = T, \end{cases} \tag{23.6}$$

where $\rho^2_{\mathrm{AO},t} = \left(1 + \sum_{i=1}^{T-t} \pi_i^2\right)^{-1}$, $\rho^2_{\mathrm{IO},t} = 1$, and $\rho^2_{\mathrm{LS},t} = \left(1 + \sum_{i=1}^{T-t} \eta_i^2\right)^{-1}$. The π_i and η_i are the coefficients of B^i in the polynomials $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots = \phi(B)/\theta(B)$ and $\eta(B) = 1 - \eta_1 B - \eta_2 B^2 - \ldots = \pi(B)/(1-B)$, respectively. The π weights are found by multiplying both sides of the definition of $\pi(B)$ by $\theta(B)$ to get $\theta(B)(1 - \pi_1 B - \pi_2 B^2 - \ldots) = \phi(B)$. For example, for an ARMA(1,1) process we have

$$\begin{aligned} 1 - \phi B &= (1 - \theta B)(1 - \pi_1 B - \pi_2 B^2 - \ldots) \\ &= 1 - (\pi_1 + \theta)B - (\pi_2 - \theta\pi_1)B^2 - (\pi_3 - \theta\pi_2)B^3 - \ldots. \end{aligned}$$

Equating the coefficients of like powers of B, we have $\pi_1 = \phi - \theta$, $\pi_2 = \theta\pi_1$, and $\pi_j = \theta\pi_{j-1} = \theta^{j-1}\pi_1$ for $j > 1$. A similar approach can also be used in calculating the η weights. One can easily verify that for an ARMA(1,1) model the corresponding η weights are $\eta_1 = \phi - \theta - 1$ and $\eta_j = \eta_{j-1} + \theta^{j-1}\pi_1$ for $j > 1$. Thus, for an AR(1) process, $\pi_j = \phi^j$ and $\eta_j = \phi - 1$ for $j \geq 1$.

Using the above results, we can construct the following test statistic for testing the existence of AO, IO, and LS at time point d:

$$\lambda_{j,d} = \frac{\hat{\omega}_{j,d}}{[\mathrm{Var}(\hat{\omega}_{j,d})]^{1/2}} = \frac{\hat{\omega}_{j,d}}{\rho_{j,d}\sigma}, \qquad j = \mathrm{AO, IO, LS}. \tag{23.7}$$

Under the null hypothesis of no outliers or level shifts, and assuming that both time d and the parameters of the ARMA model in equation (23.1) are known, the statistics $\lambda_{\mathrm{AO,t}}$, $\lambda_{\mathrm{IO,t}}$ and $\lambda_{\mathrm{LS,t}}$ are asymptotically distributed as $N(0,1)$. In practice, the time series parameters for this statistic are usually unknown and must therefore be replaced by some consistent estimates.[33,34]

To detect an AO, IO or LS at unknown position, we calculate the following statistics:

$$\lambda_{j,\max} = \max_{1\le t\le T}\left\{\left|\lambda_{j,t}\right|\right\} \qquad j = \text{AO, IO, LS.} \tag{23.8}$$

The null hypothesis of no AO, IO or LS is rejected when $\lambda_{\mathrm{j,max}}$ exceeds a pre-specified critical value. Note that it is difficult to determine the exact repeated sampling distribution of equation (23.8). Chen and Liu[35] considered a broad spectrum of π weight patterns in estimating the percentiles of (23.8) using Monte Carlo simulation. They noted that the percentiles associated with (23.6) are distinctively lower than those associated with equations (23.4) and (23.5). For an AR(1) process, for example, when $T = 200$ the estimated 1st percentile for $\lambda_{\mathrm{LS,max}}$, based on Chen and Liu's results, is between 3.3 and 3.5, while for $\lambda_{\mathrm{AO,max}}$ and $\lambda_{\mathrm{IO,max}}$ it is between 4.0 and 4.2.

In SPC, we are primarily concerned with the detection of the presence of external sources of variation manifested by changes in process level or variance. If we can determine whether the change in process behavior is due to an AO, IO or LS, we can narrow down its most probable cause. This will facilitate the quick discovery and correction of problems associated with the detected signal of process change. In the following sections, we can see that an SPC scheme based on the outliers-and-level-shift-detection statistics can be established for process monitoring purposes. Although we focus our discussion on LS detection for an AR(1) process, one can easily extend the application of the proposed procedures in effectively identifying an LS, IO, or AO under the general autogressive integrated moving average model.

23.3 BEHAVIOR OF $\lambda_{\mathrm{LS},t}$

It is known that, when an outlier or level shift occurs at some known point, the statistics (23.4), (23.5), and (23.6) are normally distributed with means ω_{AO}, ω_{IO}, and ω_{LS}, and variances $\rho^2_{\mathrm{AO},d}\sigma^2$, $\rho^2_{\mathrm{IO},d}\sigma^2$ and $\rho^2_{\mathrm{LS},d}\sigma^2$, respectively. In this section, we study the behavior of $\{\lambda_{\mathrm{LS},t} : 1 \le t \le T\}$ when a level shift or outlier occurs. Doing so may lead to ways of making the procedure described in the preceding section more sensitive to the presence of exogenous interventions or special causes when implemented on-line.

23.3.1 Effect of a level shift

In studying the effect of outliers and level shifts on $\left\{\lambda_{\mathrm{LS},t} : 1 \le t \le T\right\}$, we focus on the AR(1) model

$$Z_t - \mu = \phi(Z_{t-1} - \mu) + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{23.9}$$

where $-1 < \phi < 1$ and the random errors ε_t are i.i.d $N(0,\sigma^2)$. For the AR(1) process given by equation (23.9), $\pi(B) = 1 - \phi B$ and $\eta(B) = 1 - (\phi - 1)B - (\phi - 1)B^2 - \ldots$. Now assume that at time $t = d$ $(1 \le t \le T)$ the mean shifts from 0 to ω_{LS}. Using Tsay's notation[33], we have $Y_t = Z_t$ for $t < d$ and $Y_t = Z_t + \omega_{\mathrm{LS}}$ for $t \ge d$. It can be easily shown that

$$y_t = \begin{cases} \varepsilon_t, & t < d, \\ \omega_{\mathrm{LS}}+\varepsilon_t, & t = d, \\ (1-\phi)\omega_{\mathrm{LS}} + \varepsilon_t, & t > d. \end{cases} \tag{23.10}$$

From (23.7), we can see that $\hat{\omega}_{\mathrm{LS},t}$ is a linear combination of normally distributed random variables and is therefore normally distributed itself. The expected value and variance of $\hat{\omega}_{\mathrm{LS},t}$ for $1 \le t \le T$ can therefore be derived as follows. For $t < d$,

$$\begin{aligned} E(\hat{\omega}_{\mathrm{LS},t}) &= E\left[\rho_{\mathrm{LS},t}^2\left(y_t - \sum_{i=1}^{T-t}(\phi-1)y_{t+i}\right)\right] \\ &= \rho_{\mathrm{LS},t}^2\left[E(y_t) + (1-\phi)\left(\sum_{i=1}^{d-1}E(y_i) + E(y_d) + \sum_{i=d+1}^{T}E(y_i)\right)\right] \\ &= \rho_{\mathrm{LS},t}^2\omega_{\mathrm{LS}}(1-\phi)\left[1+(T-d)(1-\phi)\right]. \end{aligned}$$

For $t = d$,

$$\begin{aligned} E(\hat{\omega}_{\mathrm{LS},t}) &= E\left[\rho_{\mathrm{LS},d}^2\left(y_d - \sum_{i=1}^{T-d}(\phi-1)y_{d+i}\right)\right] \\ &= \rho_{\mathrm{LS},d}^2\left[E(y_d) + (1-\phi)\sum_{i=d+1}^{T}E(y_i)\right] \\ &= \omega_{\mathrm{LS}} \end{aligned}$$

Finally, for $t > d$,

$$\begin{aligned} E(\hat{\omega}_{\mathrm{LS},t}) &= E\left[\rho_{\mathrm{LS},t}^2\left(y_t - \sum_{i=1}^{T-t}(\phi-1)y_{t+i}\right)\right] \\ &= \rho_{\mathrm{LS},t}^2\left[E(y_t) + (1-\phi)\sum_{i=t+1}^{T}E(y_i)\right] \\ &= \rho_{\mathrm{LS},t}^2\omega_{\mathrm{LS}}(1-\phi)\left[1+(T-t)(1-\phi)\right]. \end{aligned}$$

Furthermore, we can show that the variance of $\hat{\omega}_{\mathrm{LS},t}$ for $1 \le t \le T$ is $\rho_{\mathrm{LS},t}^2\sigma^2$. Thus, $\{\lambda_{\mathrm{LS},t} : 1 \le t \le T\}$ has the following expected behavior:

$$E[\lambda_{\mathrm{LS},t}] = \begin{cases} \rho_{\mathrm{LS},t}\omega_{\mathrm{LS}}(1-\phi)\left[1+(T-d)(1-\phi)\right]/\sigma, & 1 \le t < d, \\ \omega_{\mathrm{LS}}/\rho_{\mathrm{LS},t}\sigma, & t = d, \\ \rho_{\mathrm{LS},t}\omega_{\mathrm{LS}}(1-\phi)\left[1+(T-t)(1-\phi)\right]/\sigma, & d < t \le T. \end{cases} \tag{23.11}$$

From (23.11), one can see that the maximum of $E[\lambda_{\mathrm{LS},t}]$ is attained at $t = d$. This corresponds to Tsay's[33] statistic for detecting level shift. It can be seen from (23.11) that when an LS occurs the expected value of $\lambda_{\mathrm{LS},t}$ becomes non-zero even for $t \neq d$. Thus, besides the magnitude of $\lambda_{\mathrm{LS},d}$, one can also explore the information given by $\lambda_{\mathrm{LS},d}$ for $t \neq d$ in detecting an LS.

To illustrate the application of the above results, we simulated an AR(1) process with $\phi = 0.9$ and $\sigma = 1$. for $t = 1, \ldots, 100$ the mean of the series is 0. Starting at time $t = 101$, we introduced an LS of magnitude $\sigma_Z = 2.254\,157$ (see Figure 23.2). Using an SCC with control limits set at $\pm 3\sigma$, we can see that the corresponding ARL is 223.30 with a standard deviation of 283.60. For the data in Figure 23.2, the change in mean was detected by the SCC at $t = 318$, or 218 readings after the change was introduced (see Figure 23.3). This indicates the poor sensitivity of the SCC.

To illustrate how the average of $\lambda_{\mathrm{LS},t}$ for $t = 1, \ldots, T$ can be used to detect a change in the level of a time series, we calculate and plot $\{\lambda_{\mathrm{LS},t} : 1 \le t \le T\}$ before and after the change occurred at $T = 100$. The calculations shown in Table 23.1 were done using standard spreadsheet software. In the table, y_1 was assumed zero, which is typical in a forecasting scenario. However, in on-line process monitoring such as that

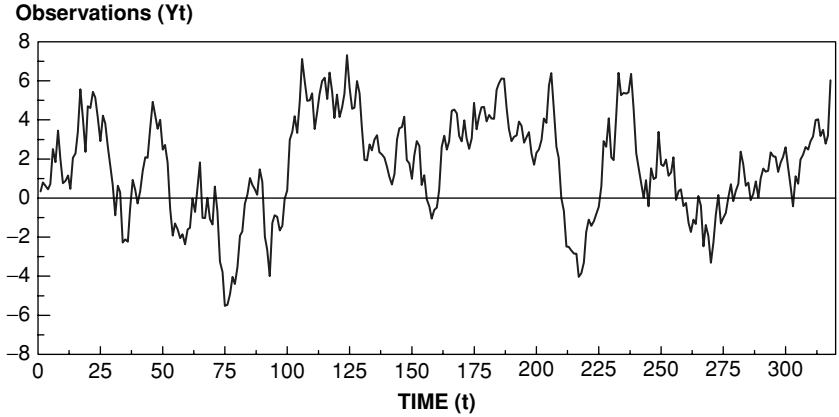


Figure 23.2 The simulated AR(1) series ($\phi = 0.9$, $\sigma^2 = 1.0$ and $\omega_{\mathrm{LS},101} = \sigma_z$).

described in the succeeding section, it is possible to determine the value of y_1 based on the preceding sets of process measurements. Since $\eta_j = \phi - 1$ $(j \geq 1)$ for an AR(1) process, the quantity $\sum_{i=1}^{T-t} \eta_i y_{t+i}$ in equation (23.6) can easily be computed by first multiplying each y_t by $\phi - 1$ (i.e. column 4) and subsequently calculating the partial sums in column 5. Knowing that $\rho^2_{\mathrm{LS},t} = [1 + (T - t)(\phi - 1)]^{-1}$, we can now calculate $\lambda_{\mathrm{LS},t}$ using the values in columns 3 and 5. The data in columns 5 and 6 are calculated for $T = 100$, while the values in columns 7 and 8 are for $T = 106$.

Figure 23.4 shows that when $T = 100$ (i.e. no change has been introduced), the average of $\{\lambda_{\mathrm{LS},t} : 1 \leq t \leq T\}$ is approximately zero. Six observations after the introduction of the LS at $t = 101$, one can readily see that the average of $\lambda_{\mathrm{LS},t}$ has shifted (see Figure 23.5). This shows the sensitivity of the average of $\lambda_{\mathrm{LS},t}$ in indicating the presence of an LS. One can therefore explore the sensitivity of the average of $\lambda_{\mathrm{LS},t}$ in designing on-line SPC monitoring schemes for detecting level shifts.

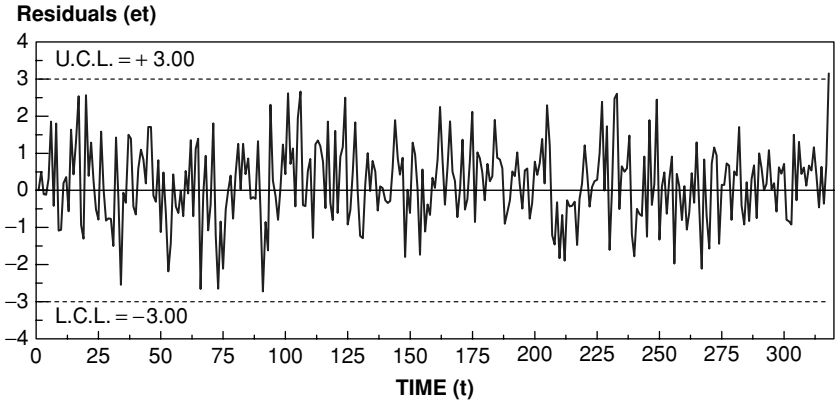


Figure 23.3 The special cause chart for the series in Figure 23.2.

Table 23.1 Calculation of $\lambda_{\mathrm{LS},t}$ using a spreadsheet.

t	Y_t	y_t	ηy_t	$\eta \sum_{i=1}^{100-t} y_{t+i}$	$\lambda_{LS,t}$ (T=100)	$\eta \sum_{i=1}^{106-t} y_{t+i}$	$\lambda_{LS,t}$ (T=106)
1	0.3434	0.0000	0.0000	−0.3669	0.2601	−1.2310	0.8598
2	0.8053	0.4962	−0.0496	−0.3173	0.5781	−1.1814	1.1746
3	0.6245	−0.1003	0.0100	−0.3273	0.1618	−1.1914	0.7658
4	0.4441	−0.1179	0.0118	−0.3391	0.1580	−1.2032	0.7636
5	0.7227	0.3230	−0.0323	−0.3068	0.4511	−1.1709	1.0538
6	2.5025	1.8520	−0.1852	−0.1216	1.4170	−0.9857	2.0066
7	1.8366	−0.4157	0.0416	−0.1632	−0.1818	−1.0273	0.4335
8	3.4523	1.7994	−0.1799	0.0167	1.2865	−0.8473	1.8809
9	2.0216	−1.0854	0.1085	−0.0918	−0.7190	−0.9559	−0.0923
10	0.7554	−1.0640	0.1064	−0.1982	−0.6281	−1.0623	−0.0012
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	0.8240	−0.5045	0.0504	0.1827	−0.6552	−0.6814	0.1643
91	−1.9793	−2.7209	0.2721	−0.0894	−2.5205	−0.9535	−1.6481
92	−2.6344	−0.8530	0.0853	−0.1747	−0.6527	−1.0388	0.1740
93	−3.9863	−1.6153	0.1615	−0.3363	−1.2365	−1.2004	−0.3903
94	−1.2763	2.3113	−0.2311	−0.1051	2.3471	−0.9692	3.0998
95	−0.8799	0.2688	−0.0269	−0.0783	0.3387	−0.9423	1.1495
96	−0.9547	−0.1628	0.0163	−0.0945	−0.0670	−0.9586	0.7587
97	−1.6546	−0.7953	0.0795	−0.1741	−0.6121	−1.0382	0.2326
98	−1.4208	0.0683	−0.0068	−0.1672	0.2332	−1.0313	1.0581
99	−0.0450	1.2337	−0.1234	−0.0439	1.2712	−0.9080	2.0704
100	0.3982	0.4387	−0.0439			−0.8641	1.2653
101	2.9764	2.6181	−0.2618			−0.6023	3.1428
102	3.3955	0.7167	−0.0717			−0.5306	1.2231
103	4.1854	1.1294	−0.1129			−0.4177	1.5244
104	3.3373	−0.4295	0.0430			−0.4606	0.0308
105	4.9489	1.9453	−0.1945			−0.2661	2.2004
106	7.1148	2.6609	−0.2661				

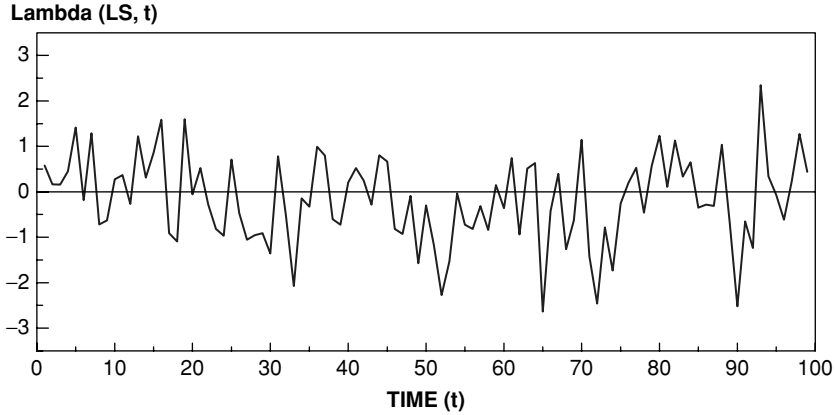


Figure 23.4 The behavior of $\lambda_{\mathrm{LS},t}$ for $t = 1, \ldots, T = 100$ (no lvel shift).

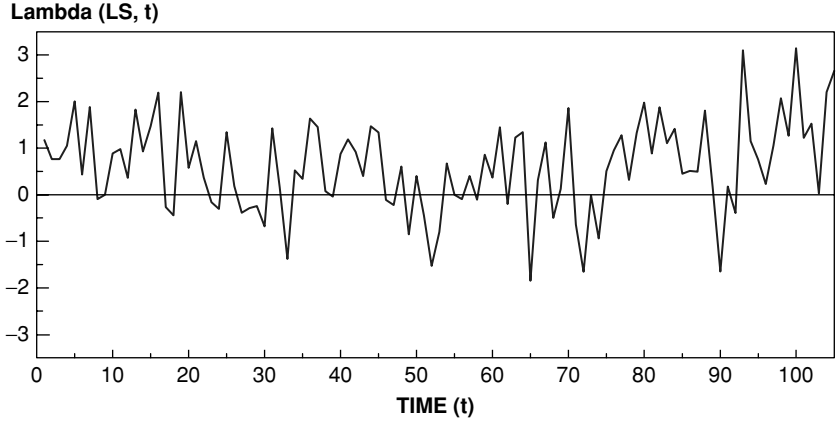


Figure 23.5 The behavior of $\lambda_{\mathrm{LS},t}$ for $t = 1, \ldots, T = 100$ (level shift of magnitude $\omega_{\mathrm{LS},101} = \sigma_z$at $t = 101$).

23.3.2 Effect of outliers

Apart from level shifts, we observed that another common manifestation of special causes of variation is the presence of additive or innovational outliers in the series. Although detecting the presence of an AO or IO is not our primary concern, it is interesting to know how $\lambda_{\mathrm{LS},t}$ behaves when these other two outliers are present. Using the approach in the preceding section, we can show that when an AO occurs between 1 and T, $\hat{\omega}_{\mathrm{LS},t}$ has the following expected behavior:

$$E(\hat{\omega}_{\mathrm{LS},t}) = \begin{cases} \rho^2_{\mathrm{LS},t}\omega_{\mathrm{AO}}(1-\phi)^2, & 1 \le t < d, \\ \rho^2_{\mathrm{LS},t}\omega_{\mathrm{AO}}\left[1-\phi(1-\phi)\right], & t = d, \\ -\rho^2_{\mathrm{LS},t}\omega_{\mathrm{AO}}\phi, & t = d+1, \\ 0, & d+1 < t \le T. \end{cases}$$

Although the first $d-1\hat{\omega}_{\mathrm{LS},t}$s are affected by $\omega_{\mathrm{AO},d}$, their magnitude is highly deflated compared with the magnitude of the AO since both $\rho^2_{\mathrm{LS},t}$ and $1-\phi$ are less than 1 for a stationary AR process. This is also the case for $t = d$. This suggests that $\hat{\omega}_{\mathrm{LS},t}$ is not suitable for detecting the presence of an AO. However, it can be seen that when $T = d$, $E(\hat{\omega}_{\mathrm{LS},d}) = E(\hat{\omega}_{\mathrm{AO},d}) = \omega_{\mathrm{AO},d}$. Thus when the AO occurs at the end of the series, it can possibly be detected by $\hat{\omega}_{\mathrm{LS},t}$ depending on the magnitude of ω_{AO}. Since this is usually the case when dealing with SPC data, control charts based on $\hat{\omega}_{\mathrm{LS},t}$ may also be effective in detecting AOs.

For an IO occurring between 1 and T, $\lambda_{\mathrm{LS},t}$ has the following expected behavior:

$$E(\hat{\omega}_{\mathrm{LS},t}) = \begin{cases} \rho^2_{\mathrm{LS},t}\omega_{\mathrm{IO}}(1-\phi), & 1 \le t < d, \\ \rho^2_{\mathrm{LS},t}\omega_{\mathrm{IO}}, & t = d, \\ 0, & d < t \le T. \end{cases}$$

The behavior of $\hat{\omega}_{\mathrm{LS},t}$ when an IO occurs is basically the same as when an AO occurs. That is, $\hat{\omega}_{\mathrm{LS},t}$ cannot suitably handle the IO when it occurs between 1 and T. Similar to

the AO case, when an IO occurs at $T = d$, $E(\hat{\omega}_{\mathrm{LS},d}) = E(\hat{\omega}_{\mathrm{IO},d}) = \omega_{\mathrm{IO},d}$. So SPC schemes based on $\hat{\omega}_{\mathrm{LS},t}$ may also detect the presence of an IO.

23.4 PROPOSED MONITORING PROCEDURE

The results in the preceding sections suggest that we can establish control charting procedures for detecting process level shifts by monitoring either $\lambda_{\mathrm{LS,\,max}}$ or $\overline{\lambda}_{\mathrm{LS},} = \sum_{t=1}^{T} \lambda_{\mathrm{LS},,t}/T$. In assessing the performance of the proposed control charting schemes, we focus on detecting the LS in an AR(1) process.

23.4.1 Control chart implementation

In implementing the proposed control charting scheme, one needs to start with an initial set of m observations taken when the process is in control. During the initial stage of implementation, the process is most likely affected by outliers and level shifts. Their presence may result in inappropriate model identification. This may subsequently lead to misidentification of process excursions. Thus, at this stage, we recommend the use of procedures that can accomplish both outlier detection and model estimation jointly.[33,35–37].

Analysis shows that control charts based on the statistic $\lambda_{\mathrm{LS},t}$ are sensitive to the choice of m. For example, simulation results show that $m = 100$ gives poor average run length (ARL) performance. A satisfactory ARL performance is achieved when $m = 200$. Note that although this m is large by conventional SPC standards, it may only correspond to a few seconds or minutes in real-time process monitoring.

As soon as the initial set of m process observations becomes available, one may start calculating $\lambda_{\mathrm{LS},t}$. In using the proposed control charting scheme, one only needs to maintain the m latest observations. That is, when a new observation becomes available, the oldest observation is discarded. Using the m latest observations, we calculate the m $\lambda_{\mathrm{LS},t}$s. To illustrate, suppose we start with an initial set of process observations $\{Z_1, Z_2, \ldots, Z_m\}$. From this initial set, we calculate the m $\lambda_{\mathrm{LS},t}$s, $\{\lambda_{\mathrm{LS},1}, \lambda_{\mathrm{LS},2}, \ldots, \lambda_{\mathrm{LS},m}\}$. When the observation Z_{m+1} becomes available, we use the window of observations $\{Z_2, Z_3, \ldots, Z_{m+1}\}$ to calculate $\{\lambda_{\mathrm{LS},2}, \lambda_{\mathrm{LS},3}, \ldots, \lambda_{\mathrm{LS},m+1}\}$. Thus, for any time period i after the initial set of m observations was taken, we have $\{Z_{i+1}, Z_{i+2}, \ldots, Z_{m+i}\}$, a window of m latest observations, from which we calculate $\{\lambda_{\mathrm{LS},i+1}, \lambda_{\mathrm{LS},i+2}, \ldots, \lambda_{\mathrm{LS},m+i}\}$. In mathematical form, we therefore need to track either

$$\lambda_{\mathrm{LS,max},i} = \max_{i+1 \le t \le m+i} \{|\lambda_{\mathrm{LS},t}|\}, \qquad i = 1, 2, 3, \ldots,$$

or

$$\overline{\lambda}_{\mathrm{LS},i} = \frac{\sum_{t=i+1}^{m+i} \lambda_{\mathrm{LS},t}}{m}, \qquad i = 1, 2, 3, \ldots.$$

The chart signals an out-of-control situation when $\lambda_{\mathrm{LS,\,max}}$ or $\overline{\lambda}_{\mathrm{LS},t}$ exceeds some prespecified control limits. These control limits can be based on the sampling distribution of $\lambda_{\mathrm{LS,max}}$ or $\overline{\lambda}_{\mathrm{LS}}$, which is difficult to obtain as these statistics are functions of dependent variables. We therefore recommend the use of simulation in determining the

appropriate control limits. This can be done by first specifying a standard or acceptable in-control ARL. A simulation program is then used to determine what control limits can be used to achieve this in-control ARL value. For the control chart based on $\lambda_{LS,max}$, the results of Chen and Liu[35] can be used as starting points for simulation.

23.4.2 Control chart performance

For an AR(1) process with $\phi > 0$, when a level shift of magnitude ω_{LS} occurrs at $t = d$, the expected value of the residual at that point in time is ω_{LS}. For $t > d$, the expected value of the residuals becomes $(1 - \phi)\omega_{LS}$. Thus, an SCC has a high chance of detecting the shift right after it occurs. After this point, the probability of an SCC detecting this shift becomes small, especially when $\phi \to 1$. An SCC is, therefore, expected to give better performance than a CUSUM chart on residuals when the change in level at time $t = d$ produces a large residual (e.g. greater than 3 in absolute value). However, given non-detection of the change in level by the SCC at $t = d$, use of the CUSUM would be recommended. Thus, both charts have advantages and disadvantages in monitoring autocorrelated processes.

The proposed control chart based on $\lambda_{LS,t}$ combines the desirable properties of the SCC and the CUSUM. It detects both abrupt and small shifts in process level. To illustrate this property of the $\lambda_{LS,\,max}$ control chart, we have chosen two possible cases of LS occurrence for an AR(1) with $\phi = 0.9$ and $\sigma = 1$. The first case is shown in Figure 23.6. Here the level change did not produce a large residual. The first 200

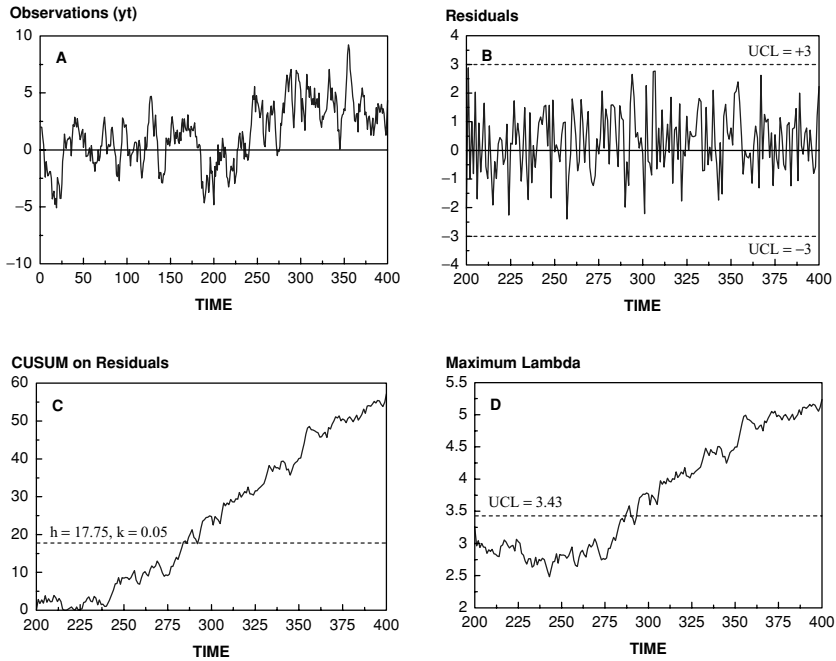


Figure 23.6 Performanace of SCC, CUSUM on residuals, and $\lambda_{LS,max}$ in detecting a level shift when the residual at the point of shift is small.

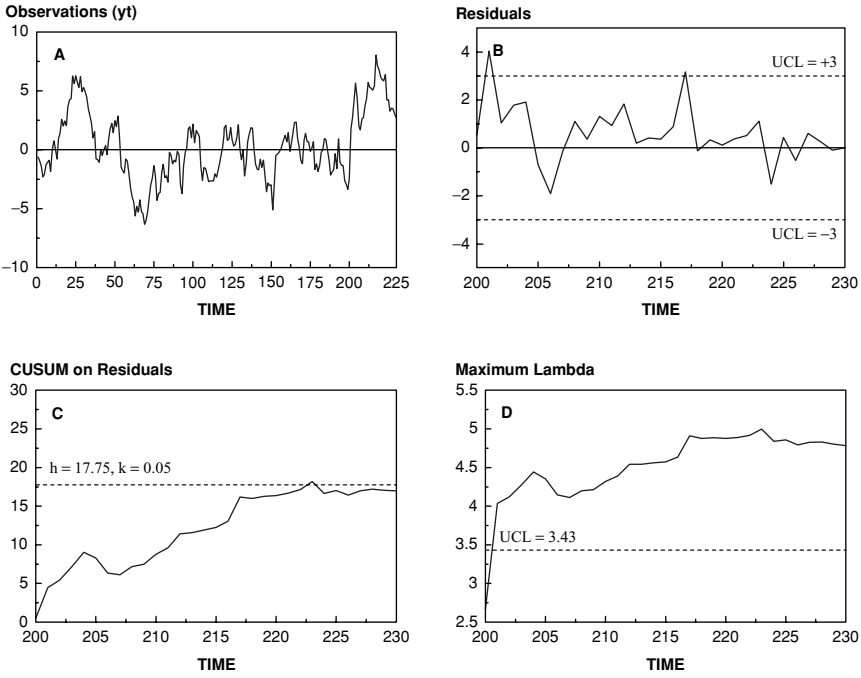


Figure 23.7 Performanace of SCC, CUSUM on residuals, and $\lambda_{LS,max}$ in detecting a level shift when the residual at the point of shift is large.

observations come from an AR(1) with $\phi = 0.9$ and mean zero. A level shift of size $1.5\sigma_Z$ was introduced at $t = 201$. Three control charts were drawn to detect change in the process level – SCC, CUSUM on residuals, and $\lambda_{LS,\,max}$.The parameters of the three control charts were chosen such that the in-control ARL is approximately 370. The SCC (Figure 23.6B) did not detect the change within the interval $201 \leq t \leq 400$. The CUSUM on residuals and the $\lambda_{LS,\,max}$ chart detected the shift at almost the same time.

The second case shows the performance of the $\lambda_{LS,\,max}$ chart when the LS results in a large residual. The time series in Figure 23.7A was produced by the same process that produced the time series in Figure 23.6A. Unlike in the previous case, one can see that the LS that occurred at $t = 201$ produced a large residual. This large residual caused the SCC to signal an out-of-control situation. Here, the CUSUM on residuals failed to quickly detect the presence of an LS. As can be seen in Figure 23.7D, the $\lambda_{LS,\,max}$ chart was able to detect the shift at t=201. Using the cases shown in Figures 23.6 and 23.7, one can expect the $\lambda_{LS,\,max}$ chart to perform better than the CUSUM on residuals or the SCC in monitoring autocorrelated processes.

23.4.3 ARL comparisons

For various levels of autocorrelation ϕ and shift δ, we compared the ARL performance of $\lambda_{LS,\,max}$ and $\bar{\lambda}_{LS}$ with the SCC using Monte Carlo simulation. Note that the shift in

mean δ is measured in terms of the standard deviation of the observations. Each simulation run is composed of 5000 iterations. The simulation program is coded in FORTRAN 77. The AR(1) process is generated using the innovation algorithm[3] while sequences of i.i.d. normal random variables are generated using the IMSL FORTRAN Library[39]. For comparison purposes, we fixed $m = 200$ and set the control limits of $\lambda_{LS,\,max}$ and $\bar{\lambda}_{LS}$ charts such that the corresponding in-control ARL is approximately equal to the in-control ARL of an SCC with $\pm 3\sigma$ limits. The comparison results are shown in Table 23.2 and illustrated in Figure 23.8. These clearly show the superiority of the control chart based on $\lambda_{LS,\,max}$. This superiority is more pronounced when $\phi > 0$. These results are not unexpected because, as shown in the preceding section, the $\lambda_{LS,\,max}$ chart possesses the desirable properties of the SCC and CUSUM on residuals. It is interesting to note that compared to $\lambda_{LS,\,max}$, $\bar{\lambda}_{LS}$ performs better in detecting small process shifts but is less sensitive in detecting large shifts particularly when ϕ is large.

23.5 CONCLUSIONS

Most traditional control charting procedures are grounded on the assumption that the process observations being monitored are independent and identically distributed. With the advent of high-speed data collection schemes, the assumption of independence is usually violated. That is, autocorrelation among measurements becomes an inherent characteristic of a stable process. This autocorrelation causes significant deterioration in control charting performance. To address this problem, several approaches for handling autocorrelated processes have been proposed. The most popular procedure utilizes either a Shewhart, CUSUM or EWMA chart of the residuals of the appropriately fitted ARMA model. However, procedures of this type possess poor sensitivity especially when dealing with positively autocorrelated processes. As an alternative, we have explored the application of the statistics used in a time series procedure for detecting outliers and level shifts in process monitoring.

The study focused on the detection of level shifts of autocorrelated processes with particular emphasis on the important AR(1) model. The results presented showed that a scheme for monitoring changes in level of autocorrelated processes can be based on either $\lambda_{LS,\,max}$ or $\bar{\lambda}_{LS}$. The $\lambda_{LS,\,max}$ control chart possesses both the desirable properties of the Shewhart and CUSUM charts. It therefore gives superior ARL performance compared with the existing procedures for detecting level shifts of autocorrelated processes. Compared to the SCC and $\lambda_{LS,\,max}$ charts, the $\bar{\lambda}_{LS}$ chart is found to be more sensitive in detecting small shifts but less sensitive in detecting large shifts.

One can easily extend the proposed control charting scheme to effectively detect the presence of additive and innovational outliers. A refined identification of the type of intervention affecting the process will allow users to effectively track the source of an out-of-control situation which is an important step in eliminating the special causes of variation. It is also important to note that the proposed procedure can also be applied when dealing with a more general autoregressive integrated moving average model.

Autocorrelated process observations mainly arise under automated data collection schemes. Such collection schemes are typically controlled by software which can be

Table 23.2 ARL comparisons (and standard deviations): SCC, $\lambda_{LS,\,max}$ and $\bar{\lambda}_{LS}$ control charts.

ϕ	Monitoring technique	Control limits	Shift (δ)							
			0.0	0.25	0.5	1.0	1.5	2.0	2.5	3.0
0.00	SCC*	±3.00	370.37	281.14	155.22	43.89	14.97	6.30	3.24	2.00
			(369.87)	(280.64)	(154.72)	(43.39)	(14.46)	(5.78)	(2.70)	(1.41)
	SCC**	±3.00	373.37	296.29	162.69	43.59	14.75	6.52	3.41	2.04
			(392.44)	(287.51)	(155.78)	(41.87)	(13.97)	(5.80)	(2.72)	(1.36)
	$\lambda_{LS,\,max}$	±3.41	372.76	96.37	31.43	9.87	5.08	3.26	2.38	1.83
			(380.16)	(82.06)	(21.91)	(5.83)	(2.71)	(1.60)	(1.09)	(0.78)
	$\bar{\lambda}_{LS}$	±1.64	377.37	64.19	30.01	14.27	9.51	7.14	5.69	4.78
			(397.93)	(46.49)	(17.57)	(7.60)	(4.74)	(3.43)	(2.67)	(2.23)
0.25	SCC*	±3.00	370.37	311.61	206.03	75.42	29.04	12.24	5.60	2.85
			(369.87)	(311.23)	(205.92)	(75.78)	(29.72)	(12.98)	(6.18)	(3.09)
	SCC**	±3.00	373.37	330.40	211.33	74.24	28.95	12.39	5.74	2.88
			(392.44)	(327.35)	(198.39)	(71.76)	(29.56)	(12.70)	(6.08)	(2.94)
	$\lambda_{LS,\,max}$	±3.44	382.00	137.41	46.65	14.49	7.20	4.32	2.90	2.03
			(393.14)	(124.29)	(34.43)	(9.34)	(4.31)	(2.57)	(1.65)	(1.11)
	$\bar{\lambda}_{LS}$	±1.63	374.80	85.80	39.66	18.37	12.02	8.98	7.03	5.88
			(396.26)	(68.30)	(25.65)	(10.24)	(6.40)	(4.60)	(3.53)	(2.86)
0.50	SCC*	±3.00	370.37	335.30	258.42	123.82	55.47	24.22	10.12	4.14
			(369.87)	(335.09)	(258.96)	(126.39)	(59.92)	(29.43)	(14.59)	(6.95)
	SCC**	±3.00	373.37	346.64	270.35	115.34	53.74	22.10	9.99	4.29
			(392.44)	(362.38)	(259.48)	(117.55)	(57.58)	(26.85)	(13.47)	(7.04)
	$\lambda_{LS,\,max}$	±3.46	381.91	189.13	72.31	22.62	10.68	5.86	3.46	2.12
			(397.66)	(185.13)	(60.71)	(16.16)	(7.39)	(4.22)	(2.64)	(1.63)
	$\bar{\lambda}_{LS}$	±1.61	378.32	123.88	54.01	24.81	15.87	11.53	9.00	7.42
			(398.68)	(111.13)	(37.46)	(14.82)	(8.91)	(6.42)	(4.96)	(4.04)
0.75	SCC*	±3.00	370.37	354.04	311.22	197.73	101.17	40.24	11.90	3.01
			(369.87)	(354.22)	(313.63)	(210.21)	(126.48)	(68.55)	(31.01)	(11.11)
	SCC**	±3.00	373.37	360.10	327.89	198.11	92.75	38.31	11.57	3.04
			(392.44)	(376.34)	(330.31)	(200.55)	(114.22)	(62.92)	(30.81)	(11.27)
	$\lambda_{LS,\,max}$	±3.46	376.78	248.75	122.06	39.56	16.53	7.25	3.06	1.46
			(387.93)	(254.92)	(113.85)	(32.90)	(15.07)	(7.83)	(3.97)	(1.62)
	$\bar{\lambda}_{LS}$	±1.50	378.08	185.39	86.39	38.91	23.88	17.29	13.14	10.50
			(398.63)	(184.40)	(70.59)	(25.95)	(15.14)	(10.69)	(8.12)	(6.51)
0.90	SCC*	±3.00	370.37	362.63	337.59	223.30	76.44	10.68	1.40	1.00
			(369.87)	(363.96)	(347.04)	(283.60)	(169.64)	(56.90)	(10.09)	(0.97)
	SCC**	±3.00	373.37	368.81	342.69	228.97	82.71	9.55	1.49	1.00
			(392.44)	(389.68)	(373.67)	(295.99)	(176.87)	(46.05)	(10.65)	(0.00)
	$\lambda_{LS,\,max}$	±3.43	374.23	298.85	183.82	60.70	15.54	2.63	1.08	1.00
			(407.55)	(334.26)	(195.79)	(70.40)	(26.24)	(6.65)	(1.03)	(0.01)
	$\bar{\lambda}_{LS}$	±1.14	381.08	268.98	150.95	65.52	38.99	26.84	20.05	14.85
			(405.41)	(278.72)	(146.98)	(52.24)	(28.76)	(19.92)	(14.87)	(11.57)

*Calculated using the program of Wardell *et al*.[2]
**Calculated using Monte Carlo simulation.

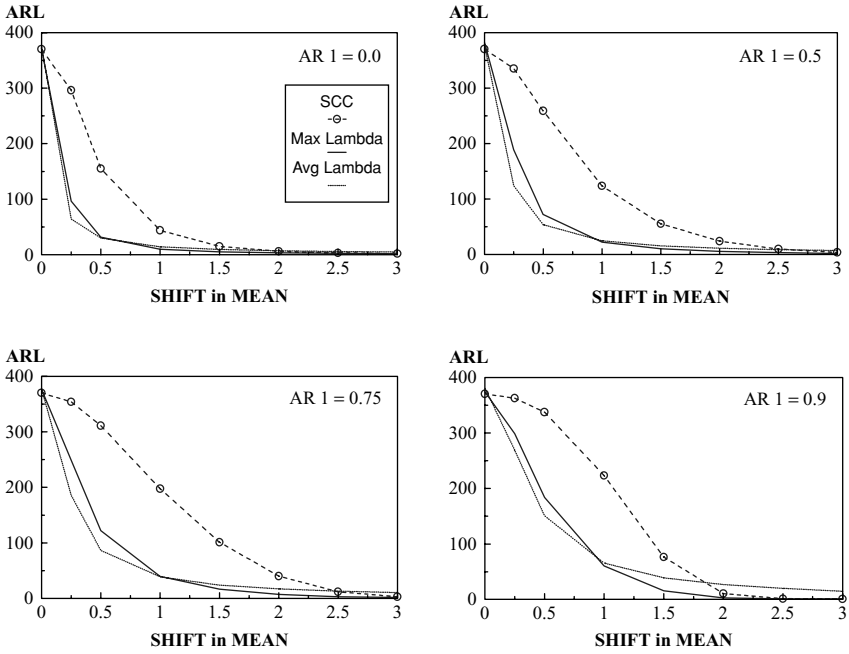


Figure 23.8 ARL comparisons: SCC, $\lambda_{\mathrm{LS,max}}$ and $\overline{\lambda}_{\mathrm{LS}}$ control charts.

upgraded to handle SPC functions. Under such an integrated scheme the usefulness of the proposed procedure will be optimized.

REFERENCES

1. Alwan, L.C. and Roberts, H.V. (1988) Time-series modelling for statistical process control. *Journal of Business and Economic Statistics*, 16, 87–95.
2. Wardell, D.G., Moskowitz, H. and Plante, R.D. (1994) Run-length distributions of special-cause control charts for correlated processes. *Technometrics*, 36, 3–17.
3. Lu, C.W. and Reynolds, M.R. (1999) EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology*, 31, 166–188.
4. Lu, C.W. and Reynolds, M.R. (1999) Control charts for monitoring the mean and variance of autocorrelated processes. *Journal of Quality Technology*, 31, 259–274.
5. Apley, D.W. and Shi, J. (1999) The GLRT for statistical process control of autocorrelated processes. *IIE Transactions*, 31, 1123–1134.
6. Apley, D.W. and Tsung, F. (2002) The autoregressive T^2 chart for monitoring univariate autocorrelated processes. *Journal of Quality Technology*, 34, 80–96.
7. Castagliola, P. and Tsung, F. (2005) Autocorrelated SPC for non-normal situations. *Quality and Reliability Engineering International*, 21, 131–161.
8. Testik, M.C. (2005) Model inadequacy and residuals control charts for autocorrelated processes. *Quality and Reliability Engineering International*, 21, 115–130.
9. Montgomery, D.C. and Mastrangelo, C.M. (1991) Some statistical process control methods for autocorrelated data (with discussion). *Journal of Quality Technology*, 23, 179–204.

10. Mastrangelo, C.M. and Brown, E.C. (2000) Shift detection properties of moving centerline control chart schemes. *Journal of Quality Technology*, 32, 67–74.
11. Alwan, L.C. (1992) Effects of autocorrelation on control chart performance. *Communications on Statistics: Theory and Methods*, 21, 1025–1049.
12. Timmer, D.H., Pignatiello, J. and Longnecker, M. (1998) The development and evaluation of CUSUM-based control charts for an AR(1) process. *IIE Transactions*, 30, 525–534.
13. Lu, C.W. and Reynolds, M.R. (2001) CUSUM charts for monitoring an autocorrelated process, *Journal of Quality Technology*, 33, 316–334.
14. Atienza, O.O., Tang, L.C. and Ang, B.W. (2002) A CUSUM scheme for autocorrelated observations. *Journal of Quality Technology*, 34, 187–199.
15. Atienza, O.O. Tang, L.C. and Ang, B.W. (1998) A SPC procedure for detecting level shifts of autocorrelated processes, *Journal of Quality Technology*, 30, 340–351.
16. English, J.R., Lee, S.C, Martin, T.W. and Tilmon, C. (2000) Detecting changes in autoregressive processes with X-bar and EWMA charts. *IIE Transactions,* 32, 1103–1113.
17. Loredo, E.N, Jearkpaporn, D. and Borror, C. (2002) Model-based control chart for autoregressive and correlated data. *Quality and Reliability Engineering International,* 18, 489–496.
18. Alwan, L.C and Radson, D.(1992) Time-series investigation of subsample mean charts. IIE Transactions, 24(5), 66–80.
19. Krieger, C.A. Champ, C.W. and Alwan, L.C. (1992) Monitoring an autocorrelated process. Presented at the Pittsburgh Conference on Modeling and Simulation.
20. Atienza, O.O., Tang, L.C. and Ang B.W. (2002) Simultaneous monitoring of sample and group autocorrelations. *Quality Engineering*, 14, 489–499.
21. Dyer, J.N., Conerly, D.M. and Adams, B.M. (2003) A simulation study and evaluation of multivariate forecast based control charts applied to ARMA processes. *Journal of Statistical Computation and Simulation*, 73, 709–724.
22. Runger, G.C. and Willemain, T.R. (1996) Batch-means control charts for autocorrelated data. *IIE Transactions*, 28, 483–487.
23. Sun, J. and Xu, L. (2004) Batch average control chart, *ASQ Annual Quality Congress Proceedings*, 2004; p. 58.
24. Balkin, S.D. and Lin, D.K. (2001) Performance of sensitizing rules on Shewhart control charts with autocorrelated data. *International Journal of Reliability, Quality and Safety Engineering*, 8, 159–171.
25. Zhang, N.F. (2000) Statistical control charts for monitoring the mean of a stationary process. *Journal of Statistical Computation and Simulation*, 66, 249–258.
26. Alwan, A.J. and Alwan, L.C. (1994) Monitoring autocorrelated processes using mutivariate quality control charts. *Proceedings of the Decision Sciences Institute Annual Meeting*, 3, 2106–2108.
27. Young, T.M. and Winistorfer, P.M. (2001) The effects of autocorrelation on real-time statistical process control with solution for forest products manufacturers. *Forest products Journal*, 51, 70–77.
28. Runger, G.C. (1996) Multivariate statistical process control for autocorrelated processes. *International Journal of production Research*, 34, 1715–1724.
29. Montgomery, D.C. (2005) Introduction to Statistical Quality Control, 5th edn. Hoboken, NJ: John Wiley & Sons, Inc.
30. Noorossana, R. and Vaghefi, S.J.M. (2005) Effect of autocorrelation on performance of the MCUSUM control chart. *Quality and Reliability Engineering International*, 22, 191–197.
31. Kalgonda, A.A. and Kulkarni, S.R. (2004) Multivariate quality control chart for autocorrelated processes. *Journal of Applied Statistics*, 31, 317–327.
32. Box, G.E.P., Jenkins G.M. and Reinsel G.C. (1994) *Time Series Analysis, Forecasting and Control*, Englewood Cliffs, NJ: Prentice Hall. Cliffs, NJ.
33. Tsay, R.S. (1988) Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7, 1–20.
34. Pankratz A. (1991) *Forecasting with Dynamic Regression Models*, John Wiley, New York: John Wiley & Sons, Inc.

35. Chen, C. and Liu, L.M. (1993) Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88, 284–297.
36. Chang, I., Tiao, G. and Chen C. (1988), Estimation of time series paremeters in the presence of outliers. *Technometrics*, 30, 193–204.
37. Balke, N.S. (1993) Detecting level shifts in time series. *Journal of Business and Economic Statistics*, 11, 81–92.
38. Brockwell, P.J. and Davis, R.A. (1987) *Time Series Analysis: Theory and Methods*. New York: Springer-Verlag.
39. IMSL(1991) *User's Manual Statistical Library,* Version 2. Houston, TX: IMSL.

24

Cumulative Sum Charts with Fast Initial Response

L. C. Tang and O. O. Atienza

In this chapter, we give a short review of the two types of cumulative sum (CUSUM) chart, namely the tabular CUSUM and V-mask CUSUM, and highlight their head-start features, also called fast initial response (FIR) for CUSUM. A more detailed exploration of FIR for V-mask CUSUM is presented as this is its first appearance in a textbook. Its underlying working principle is explained and two variants are presented.

24.1 INTRODUCTION

In conventional Shewhart control charts, only information contained in the last sample observation is used in process monitoring. As a result they are insensitive to small (about 1.5σ or less) but persistent shifts. To mitigate this deficiency, sensitizing rules, also called 'supplementary run rules', have been used in conjunction with Shewhart charts. However, Champ and Woodall[1] showed that, even with such run rules, the Shewhart chart is not as sensitive as the cumulative sum (CUSUM) control chart in detecting small, sustained shifts in the process mean.

The CUSUM chart was first introduced by Page[2], and has been studied by many authors.[3–5] In this chapter, the application of the CUSUM chart for monitoring changes in the process mean with fast initial response (FIR) is presented. The results can be extended to enable CUSUM charts to be used to monitor other parameters. To detect shifts in the process mean, the CUSUM chart is implemented by summing deviations from the nominal or target value. When this sum exceeds a specified threshold value, one concludes that there has been a change in process mean. There are two ways to represent CUSUMs, the tabular (sometimes called algorithmic) CUSUM, and the V-mask, which is the original form of the CUSUM.

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

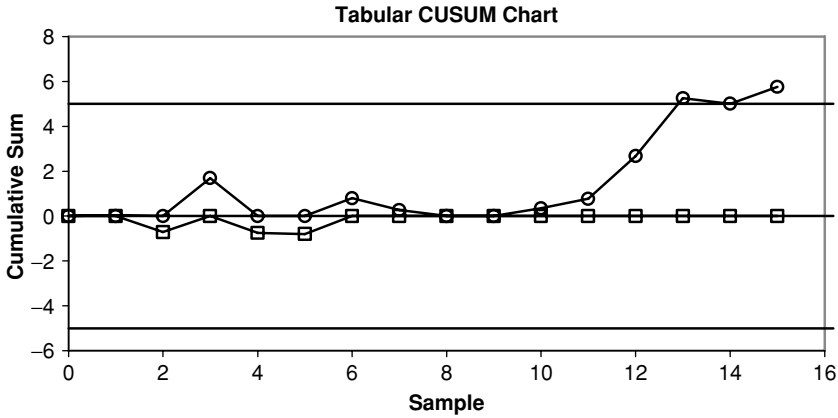


Figure 24.1 Tabular CUSUM.

24.1.1 Tabular CUSUM

Let us assume, without loss of generality, that we are monitoring a product or process characteristic that, in the absence of special causes of variation, is independent and normally distributed with mean $\mu_0 = 0$ and variance $\sigma^2 = 1$. The tabular CUSUM scheme is given by

$$\text{CUSUM}_i^+ = \max\left[0, \text{CUSUM}_{i-1}^+ + y_i - k\right],$$
$$\text{CUSUM}_i^- = \max\left[0, \text{CUSUM}_{i-1}^- - y_i - k\right],$$

where y_i is the ith observation from the process.[6] The parameter k is usually called the reference value and is determined by the shift in mean level which the CUSUM chart is designed to detect. Bissel[7] recommended a reference value of $k = \Delta/2$, where Δ is the smallest shift in mean as a multiple of σ to be detected. The constant CUSUM_0^+ (or CUSUM_0^-) is traditionally set to zero except when FIR is intended (see below). A typical tabular CUSUM is shown in Figure 24.1.

This CUSUM scheme signals an out-of-control situation as soon as CUSUM_i^+ or CUSUM_i^- exceeds the threshold value h, which is usually set to 4 or 5 for a reference value of $k = 1/2$.[8] This is recommended based on desirable average run length (ARL) properties against a shift of about 1 standard deviation in the process mean.

The action taken following an out-of-control signal on a CUSUM scheme is identical to that for other control schemes; one should search for the assignable cause, take any corrective action required, and then reinitialize the CUSUM to zero.

24.1.2 V-mask CUSUM

For a two-sided scheme, one monitors the quantity

$$\text{CUSUM}_j = \sum_{i=1}^{j} y_i, \quad j = 1, 2, \ldots, n,$$

where y_i is the ith sample reading with zero mean and unit variance and n represents the number of samples. At any time point $i = n$, we have n CUSUMs. These may fluctuate about the horizontal direction $\mu_0 = 0$ even when the process is on target.

Barnard[9] showed that a significant change in mean of the y_i can be detected by using a V-mask. The V-mask is determined by two parameters: the lead distance d and the angle θ of the mask arm with respect to the horizontal axis. Johnson[10] proposed a procedure for determining the value of d and θ using the theory of sequential probability ratio test (SPRT):

$$d = \frac{2}{\Delta^2} 1\text{n}\left(\frac{1-\beta}{\alpha/2}\right), \tag{24.3}$$

$$\tan\theta = \frac{\Delta}{2A},$$

where Δ is the amount of shift to be detected (as a multiple of σ), A is a CUSUM chart scale factor, α is the type I error or the risk of false alarm, and β is the type II error or the risk of failing to raise an alarm. Johnson's approach is still the most commonly used procedure for designing a V-mask. Figure 24.2 shows a typical V-mask with its parameters.

24.1.3 Equivalence of tabular and V-mask CUSUM

The two representations of CUSUM are equivalent if $k = A\tan\theta$ and $h = Ad\tan\theta = dk$, where A is the scale factor which represents the ratio between the vertical scale

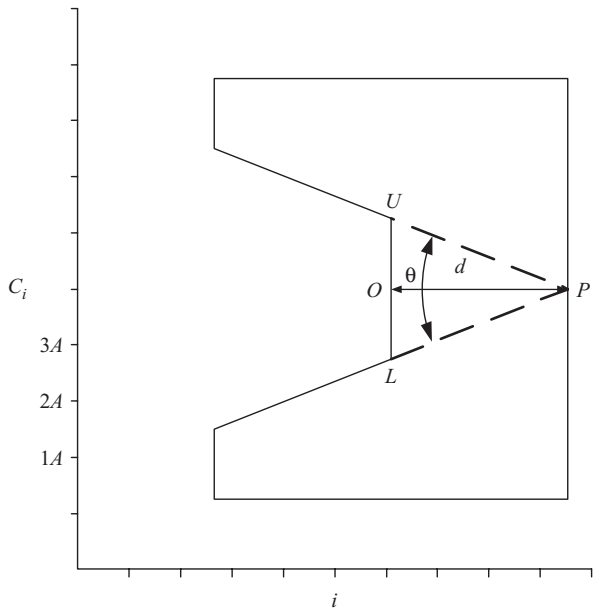


Figure 24.2 A V-mask and its parameters.

for each unit of the distance along the x-axis between consecutive points plotted in the V-mask scheme (see Figure 24.2).

24.2 FAST INITIAL RESPONSE

During the start-up stage, CUSUM_0^+ and CUSUM_0^- are given a value of zero. However, Lucas and Crosier[11] showed that the CUSUM may not reset to zero during start-up or after an out-of-control situation is detected. This non-zero start-up for CUSUM is called the fast initial response (FIR).

24.2.1 FIR for tabular CUSUM

The FIR (tabular) CUSUM for upward/downward shift is defined by

$$\begin{aligned} \text{CUSUM}_i^+ &= \max\left[0, \text{CUSUM}_{i-1}^+ + y_i - k\right], \qquad \text{CUSUM}_0^+ = H^+, \\ \text{CUSUM}_i^- &= \max\left[0, \text{CUSUM}_{i-1}^- - y_i - k\right], \qquad \text{CUSUM}_0^- = H^-, \end{aligned} \tag{24.4}$$

and signals if $\text{CUSUM}_i^+ > h$ or $\text{CUSUM}_i^- > h$.

The quantities H^+ and H^- are the FIR. Lucas and Crosier[11] recommended an FIR value of $h/2$. The motivation for the FIR CUSUM is that, if the process starts in an out-of-control or off-target state, then starting the CUSUM part way toward the threshold value h will hasten the out-of-control signal. If, however, the process is not out of control, then it is likely that the CUSUM will soon drop back to zero, after which the FIR CUSUM behaves like a conventional zero-start CUSUM. Thus, it has little effect when the process is in control or on target.

24.2.2 FIR for V-mask CUSUM

While the head-start feature is well known for tabular CUSUM, its counterpart for the V-mask CUSUM was not available until 1999 the work of Atienza *et al.*[12] As a result, standard references[5,8] do not cover the FIR feature for the V-mask CUSUM. Nevertheless, practitioners seem to prefer the ease of interpretation and the visual effect of V-mask CUSUM. It is thus instructive to give a detailed presentation of the FIR feature for the V-mask CUSUM.

In order to implement the FIR feature in the V-mask scheme, we need to understand the importance of the point at the origin (0,0) which represents a CUSUM value of when $i = 0$. In a tabular CUSUM without an FIR, the first observation y_1 signals an out-of-control situation when its value exceeds $h + k$. Under the V-mask scheme, this situation is represented by the point at the origin (0,0) that is outside the arms of the mask. The origin (0,0) is thus the key consideration for head start feature during the start-up stage. In using the tabular CUSUM with FIR, the first observation y_1 signals an out-of-control situation when its value exceeds $h + k -$ FIR. In the V-mask scheme, this is precisely indicated by (0,0) outside the arms of the mask with the first point plotted being $\text{CUSUM}_1^+ = y_1 +$ FIR. In implementing the FIR feature under the

V-mask scheme, one may think of using (0, FIR) as the first point to be plotted in the chart and the next point being $\text{CUSUM}_1^+ = y_1 + \text{FIR}$. This is only equivalent to shifting the CUSUM plot by a magnitude given by the FIR. Thus, in using this approach, we can detect a signal only when CUSUM_1^+ exceeds $h + k$, which is precisely equivalent to not implementing an FIR. By analogy, the FIR value must be added only to the first process observation and the point at the origin (0,0) still represents a point that must be within the arms of the mask when the process is in control. Thus in implementing the FIR feature under the V-mask scheme, it is only the first CUSUM value that is altered, that is, by adding or subtracting the FIR value. This means that for a two-sided monitoring scheme we must calculate the following CUSUMs:

$$\text{CUSUM}_i^+ = \begin{cases} 0, & i = 0, \\ y_1 + \text{FIR}, & i = 1, \\ \sum_{j=1}^{i} y_i + \text{FIR}, & i > 1, \end{cases} \tag{24.5}$$

and

$$\text{CUSUM}_i^- = \begin{cases} 0, & i = 0, \\ y_1 - \text{FIR}, & i = 1, \\ \sum_{j=1}^{i} y_i - \text{FIR}, & i > 1. \end{cases} \tag{24.6}$$

Since equation (24.5), which adds the FIR value to the first observation, is meant to rapidly detect a process mean above the target value, we must use the lower arm of the mask for CUSUM_i^+. Similarly, for CUSUM_i^-, we must use the upper portion of the mask.

24.2.2.1 Numerical Illustration

Here, by way illustration, the one-side CUSUM example given by Lucas and Crosier[11] is replicated. Table 24.1 gives the initially-out-of-control case, in which the mean of the observations is 1.36 while the desired mean is zero.

Using the data in Table 24.1, we detected the out-of-control condition at time $i = 7$ when the tabular CUSUM value ($k = 0.5$) exceeded the decision interval, $h = 5$. Similarly, at $i = 7$, the V-mask CUSUM depicts points that lie outside the mask arms (see Figure 24.3).

On the other hand, by using the FIR feature, one can expect a speedier detection of change when the process starts in an out-of-control state. For the data in Table 24.1, the tabular FIR CUSUM detects the out-of-control condition one time period earlier (i.e. at $i = 3$). This tabular FIR CUSUM can also be represented in V-mask form using equations (24.5) and (24.6). Both the tabular and the V-mask form of the FIR CUSUM are displayed in Figure 24.4. In both FIR schemes, we use the control chart parameters given by Lucas and Crosier:[11] $h = 5.0$, $k = 0.5$, and the recommended FIR value of $h/2 = 2.5$.

Table 24.1 FIR CUSUM example: initially-out-of-control case; $h = 5$, $k = 0.5$ (from Table 5, Lucas and Crosier[11]).

		$CUSUM_i^+$			
		Tabular		V-mask	
i		No FIR	FIR	No FIR	FIR
0	–	0.0	2.5	0	0
1	0.8	0.3	2.8	0.8	3.3
2	1.9	1.7	4.2	2.7	5.2
3	1.4	2.6	5.1**	4.1	6.6
4	2.0	4.1	6.6**	6.1	8.6
5	1.1	4.7	7.2**	7.2	9.7
6	0.7	4.9	7.4**	7.9	10.4
7	2.6	7.0*	9.5**	10.5	13.0

*CUSUM out-of-control signal.
**FIR CUSUM out-of-control signal.

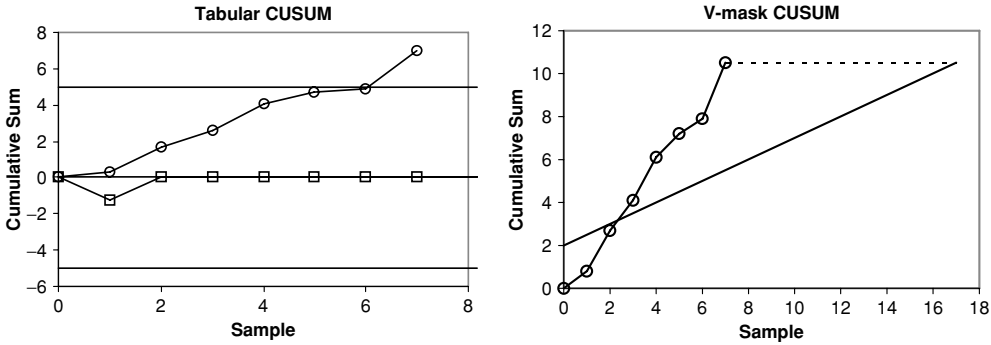


Figure 24.3 The Tabular and V-mask CUSUM without FIR for the data in Table 24.1.

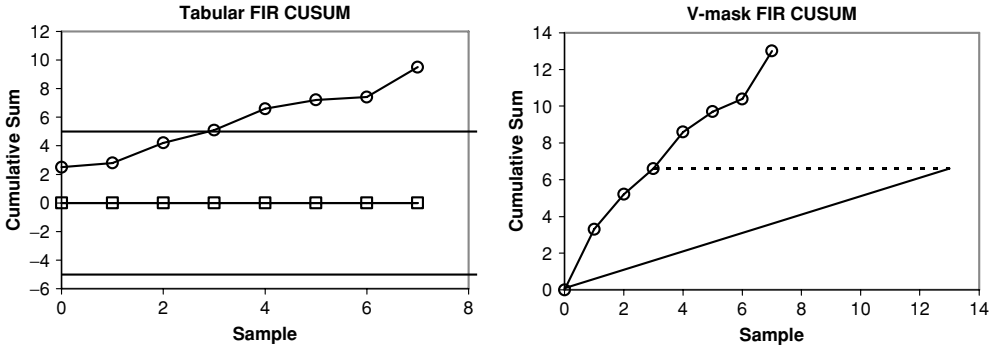


Figure 24.4 The tabular and V-mask CUSUM without FIR for the data in Table 24.1.

24.2.3 CUSUM^n

From the above it is clear that if the FIR V-mask scheme is used for a two-sided test, two charts are needed: one for checking for an increase and another one for a decrease in the mean. As an alternative, Atienza *et al*.[12] proposed the CUSUM^n, to combine the two charts into one.

Note that equations (24.5) and (24.6) differ only by the FIR at $i = 1$. By subtracting the FIR from (24.5) and adding the same value to (24.6), we have

$$\text{CUSUM}_i^n = \begin{cases} c_0, & i = 0, \\ \sum_{j=1}^{i} x_i, & i > 0, \end{cases} \tag{24.7}$$

where

$$c_0 = \begin{cases} -\text{FIR}, & \text{for CUSUM}_0^+, \\ \text{FIR}, & \text{for CUSUM}_0^-. \end{cases} \tag{24.8}$$

Thus, by using (24.7) and (24.8), we can use the typical CUSUM mask and an out-of-control situation will be signaled when c_0 is outside the mask arms. Atienza *et al*.[12] showed that (with simulated data), this procedure is applicable to all CUSUM mask forms.

As an illustration, the data in Table 24.1 are plotted in Figure 24.5. The V-mask FIR CUSUM shown in Figure 24.5 is essentially the same as the typical V-mask CUSUM except that the point at the origin (0,0) which is also used in interpreting the CUSUM is replaced by two points: (0,−FIR) and (0,+FIR). Thus the V-mask FIR CUSUM looks like a fork at the start-up stage. However, one must be aware that this alternative form of the V-mask signals only when $i = 4$, while the two-chart V-mask in Figure 24.4 signals when $i = 3$. This is due to the fact that the effect of the FIR in this alternative

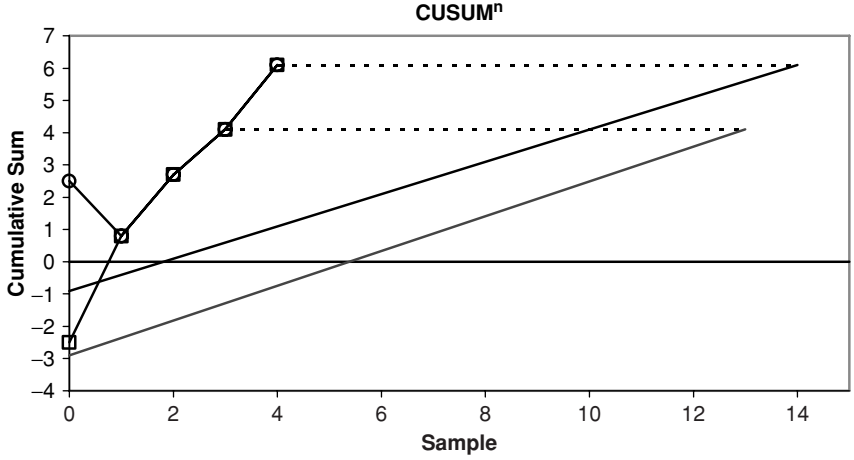


Figure 24.5 The alternative form of the V-mask CUSUM with FIR.

Table 24.2 Comparison of ARL for the tabular and V-mask FIR CUSUM, $k = 0.5$ (from Wardell *et al.*[11] and Table 2, Atienza *et al.*[12]).

Mean n shift (σ)	$h = 4$						$h = 5$					
	FIR=1		FIR=2		FIR=3		FIR=1.25		FIR=2.50		FIR=3.75	
	Tab.	V-mask k	Tab.	V-mask k	Tab.	V-mask k	Tab.	V-mask k	Tab.	V-mask k	Tab.	V-mask k
0.00	163	161.27	149	145.95	106	105.40	459	460.24	430	432.10	324	326.95
0.25	71.1	72.52	62.7	63.26	43.4	41.68	135	137.07	122	124.05	87.1	86.21
0.50	24.4	24.40	20.1	20.17	13.1	12.93	34.9	35.23	28.7	28.84	18.5	18.28
0.75	11.6	11.43	8.97	9.15	5.67	5.77	14.8	14.69	11.2	11.01	6.89	6.80
1.00	7.04	7.03	5.29	5.15	3.36	3.35	8.62	8.53	6.35	6.35	3.88	3.93
1.50	3.85	3.87	2.86	2.86	1.90	1.91	4.61	4.63	3.37	3.39	2.17	2.10
2.00	2.70	2.69	2.01	2.01	1.40	1.38	3.20	3.22	2.36	2.37	1.54	1.53
2.50	2.12	2.13	1.59	1.56	1.18	1.18	2.50	2.51	1.86	1.86	1.26	1.37
3.00	1.77	1.78	1.32	1.33	1.07	1.07	2.10	2.11	1.54	1.54	1.11	1.10
4.00	1.31	1.32	1.07	1.07	1.00	1.00	1.61	1.60	1.16	1.16	1.01	1.01

is only limited to the origin point whereas the two-chart V-mask's FIR is added into the consecutive CUSUM.

24.2.4 ARL comparisons

Using the Markov chain approach, Lucas and Crosier[11] analyzed the average run length properties of the tabular FIR CUSUM scheme. Similarly, by comparing ARL properties, Atienza *et al*.[12] further established the equivalence of the tabular FIR CUSUM and the V-mask FIR CUSUM. The ARL values for the tabular FIR CUSUM are given by Lucas and Crosier,[11] while those for the V-mask are obtained by Monte Carlo simulation. The results of the comparison by Atienza *et al.*[12] are replicated here in Table 24.2. It is clear that, in addition to the analysis concerning the tabular and V-mask FIR CUSUMs, the ARL comparison in Table 24.2 further supports the equivalence of the two FIR CUSUM schemes.

24.3 CONCLUSIONS

The FIR feature is found to be useful in implementing the CUSUM chart in its tabular form. In this chapter, we have discussed a procedure for developing an equivalent scheme to the V-mask CUSUM. The V-mask FIR CUSUM requires only a simple modification of the first point in calculating the CUSUM but requires plotting two separate charts. An alternative CUSUM, CUSUM^n, is thus proposed to overcome the need for this. That is, instead of using the point at the origin (0,0), two points are plotted, namely (0,−FIR) and (0,+FIR). This makes the V-mask FIR CUSUM looks like a fork at the start-up stage. This procedure of implementing the FIR feature in V-mask form is also applicable to all other forms of CUSUM masks.

REFERENCES

1. Champ, C.W. and Woodall, W.H. (1987) Exact results for Shewhart control charts with supplementary run rules. *Technometrics,* 29, 393–399.
2. Page, E.S. (1954) Continuous inspection schemes. *Biometrika*, 41, 100–115.
3. Gan, F.F. (1991) An optimal design of CUSUM quality control charts. *Journal of Quality Technology*, 23, 278–286.
4. Lucas, J.M. (1976) The design and use of V-mask control schemes. *Journal of Quality Technology*, 8, 1–12.
5. Hawkins, D.M. and Owell, D.H. (1998) *Cumulative Sum Charts and Charting for Quality Improvement*. New York: Springer-Verlag.
6. Kemp, K.W. (1961) The average run length of the cumulative sum chart when a V-mask is used. *Journal of the Royal Statistical Society* B, 23, 149–153.
7. Bissel, A.F. (1969) CUSUM techniques for quality control (with discussion). *Applied Statistics,* 18, 1–30.
8. Montgomery, D.C. (2005) *Introduction to Statistical Quality Control*, 5th edition. Hoboken, NJ: John Wiley & Sons Inc.
9. Barnard, G.A. (1959) Control charts and stochastic processes. *Journal of the Royal Statistical Society* B, 21, 239–257.

10. Johnson, N.L. (1961) A simple theoretical approach to cumulative sum control charts. *Journal of the American Statistical Association*, 54, 835–840.
11. Lucas, J.M. and Crosier, R.B. (1982) Fast initial response for CUSUM quality control schemes: Give your CUSUM a head start. *Technometrics*, 24, 199–205.
12. Atienza, O.O., Tang, L.C. and Ang, B.W. (1999) Fast initial response in CUSUM mask scheme. *Quality Engineering*, 11, 541–546.

25

CUSUM and Backward CUSUM for Autocorrelated Observations

L. C. Tang and O. O. Atienza

In this chapter we present a backward CUSUM (BCUSUM) scheme based on ideas from the forecasting literature. A uniformly most powerful test for detecting level shifts in the mean is derived for this scheme. An equivalent CUSUM scheme based on its mirror image is shown to have to a parabolic control boundary. The proposed scheme is further expounded for monitoring autocorrelated data. The parameter for determining the control limits can be selected based on the desired ARL. Examples are given to illustrate the idea and the application of these schemes.

25.1 INTRODUCTION

One disadvantage in using the V-mask CUSUM scheme or its tabular counterpart (see chapter 24) is that it is insensitive to large process changes. Lucas[1] noted that this is related to the anomaly of Wald's sequential likelihood ratio test on which the CUSUM scheme is based. He observed that the V-mask constructed based on Wald's test is not uniform against other alternatives. This prompted him to modify the V-mask and suggest a semiparabolic mask. Independently, Bissell[2] proposed a similar semiparabolic mask. Rowlands *et al.*,[3] on the other hand, proposed a snub-nosed V-mask which comprises superimpositions of several V-masks. These modifications were done to make the CUSUM more uniform in detecting shifts in the mean of normal random variables.

In designing a control rule for CUSUM, the purely parabolic mask has always been mentioned as a candidate. For example, Barnard[4] noted that some of his colleagues were using the parabolic mask but gave no further details. Van Dobben de Bruyn[5] also noted that there were reasons to believe that segments of parabolas would be

Six Sigma: Advanced Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap
© 2006 John Wiley & Sons, Ltd

even better than the V-mask but gave no analysis of the properties of such tests. More recently, Wiklund[6] described a parabolic CUSUM based on Lorden's results.[7]

On the other hand, with the widespread use of high-speed data collection schemes, process measurements are often serially correlated even when there are no special causes of variation affecting the system. The presence of autocorrelation violates the independence assumption in many statistical process control (SPC) charting procedures and degrades the performance of classical SPC procedures.[8,9] When the process measurements are serially correlated, a popular monitoring approach is to model the inherent autocorrelation using the Box–Jenkins autoregressive moving average (ARMA) model

$$\phi_p(B)y_t = \phi_0 + \theta_q(B)\varepsilon_t \tag{25.1}$$

where $\{y_t\}$ is a stationary time series representing the process measurements, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is an autoregressive polynomial of order p, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is a moving average polynomial of order q, B is the backshift operator, and $\{\varepsilon_t\}$ is a sequence of normally and independently distributed random errors with mean zero and constant variance σ_ε^2. If we let $\hat{y}_t$ represent the predicted value from an appropriately identified ARMA model, then the sequence of residuals $\{e_1 = y_1 - \hat{y}_1, e_2 = y_2 - \hat{y}_2, \ldots\}$ will behave like an independent and identically distributed (i.i.d.) random variable.[10] In detecting process changes, classical SPC charts can therefore be applied to the residuals of the appropriately chosen ARMA model. However, the performance of these control charts is very different from the i.i.d. case.[11]

Wardell *et al*.[11] showed that the Shewhart chart on residuals (SCR) has poor sensitivity when the process is highly positively autocorrelated. To address this issue, Runger *et al*.[12] suggested the use of CUSUM in monitoring the residuals (CUSUMR). Yaschin[13] proposed a transformation that generates an i.i.d. sequence of random variable when the process is in control. He suggested monitoring the resulting transformed series using a CUSUM chart. Wardell *et al*.[14] explained that Yaschin's approach is essentially equivalent to monitoring the residuals using a CUSUM chart.

In this chapter, a purely parabolic mask for CUSUM, proposed by Atienza *et al*.,[15] is presented. This mask, which is based on a uniformly most powerful (UMP) test, is motivated by constructing the dual relationship of CUSUM and backward CUSUM (BCUSUM). In addition, an alternative CUSUM scheme that directly utilizes the autocorrelated observations in monitoring changes in process mean is presented.

25.2 BACKWARD CUSUM

The BCUSUM scheme was originally developed by Harrison and Davies[16] for routine forecast monitoring. It is based on the idea that if we can guess the time j at which the mean of the forecast errors $(e_1, e_2, \ldots)$ shifted, then the sum or the average of $\{e_t\}$ since the change occurred would be the best indicator of the presence of a special cause. In practice, the time j when the shift in mean occurred is unknown a priori. One way to overcome this problem is to calculate a series of partial sums of all the past errors as

follows (Harrison and Davies [16]):

$$S_1 = e_t,$$

$$S_2 = e_t + e_{t-1},$$

$$S_3 = e_t + e_{t-1} + e_{t-2}, \text{ etc.}$$

These quantities are called backward cumulative sums or BCUSUMs. In applying this scheme, if we have, say, n observations or forecast errors, we need to calculate n sums. The monitoring is done on a real-time basis, thus n increases as the number of observations increases. In a typical forecasting activity, we are only interested in the first few values of S. Thus, most authors, particularly those dealing with short-term forecasting, suggest the use of the following BCUSUMs:[17,18]

$$\begin{aligned} S_{t,1} &= e_t, && t = 1, 2, \ldots, m, \\ S_{t,2} &= e_t + e_{t-1}, && t = 2, 3, \ldots, m, \\ &\ldots \\ S_{t,m} &= e_t + e_{t-1} + \cdots + e_{t-m+1}, && t = m, \end{aligned}$$

where $e_1, e_2, \ldots, e_m$ represent the latest m forecast errors being tracked. To determine whether the system is in control, Harrison and Davies[16] proposed establishing the control limits $L_1, L_2, \ldots, L_6$ for the last six BCUSUMs (i.e., $S_{t,1}, S_{t,2}, \ldots, S_{t,6}$). As long as these are within the specified control limits, the system is deemed in control or stable. Since the variance of the partial sums increases as the number of errors being summed increases, we can expect that $L_1 < L_2 < \cdots < L_6$. Harrison and Davies[16] suggested the use of the following equation for calculating the control limits:

$$L_i = \sigma_\varepsilon w(i + h),$$

where σ_ε represents the standard deviation of the errors, and w and h are parameters to be chosen. One typically selects the values of w and h using simulation.

To illustrate how a BCUSUM works, we use the example given by Gardner.[19] For this data set, the parameters $\sigma_\varepsilon = 10$, $w = 1$, and $h = 2$ were used to establish the control limits. In the example below, the change in mean of the forecast errors was detected in period 6 when $S_{6,2}$ exceeded L_2.

Recognizing that forecast monitoring is done sequentially, one can easily see that Table 25.1 contains unnecessary calculations. The only important information for control

Table 25.1 Example of the BCUSUM method.[19]

Period	Forecast error	S_1	S_2	S_3	S_4	S_5	S_6
1	−10	−10					
2	20	20	10				
3	15	15	35	25			
4	5	5	20	40	30		
5	−25	−25	−20	−5	15	5	
6	−25	−25	−50	−45	−30	−10	−20
Control limits: L_1 **to** L_6		±30	±40	±50	±60	±70	±80

purposes is that contained in the last row. The BCUSUMs in rows 1–5 are of no use at all unless we are trying to detect change not on an on-line or real-time basis. For example, it is easy to see that when there is a BCUSUM value in row 4 that exceeds the specified control limits, one does not need the ensuing observations (i.e., observations 5 and 6) to conclude that there has been a change in mean of the forecast errors. At period 4, the forecaster already knows that there has been a change or intervention affecting the system. Thus, if one needs to implement a BCUSUM considering only the last few forecast errors, one need only calculate the following BCUSUM:

$$S_i = \sum_{j=1}^{i} e_{n-j+1}, \quad i = 1, 2, \ldots, m,$$

where m represents the number of latest forecast errors being monitored.

One drawback in basing a monitoring scheme only on the latest few m observations is that small process shifts will be difficult to detect. In detecting a change in the mean of a normal random variable, for example, the ARL of a two-sided CUSUM chart ($h = 4.8, k = 0.5, ARL_0 = 390$) for a 0.25σ mean shift is approximately 120. Monitoring only the latest 6–10 observations will not provide enough sensitivity to detect such a change. This ARL figure may not have any practical importance in forecasting since forecasters are usually dealing with few samples. However, as pointed out by Atienza *et al.*,[20] forecast monitoring techniques are also useful in process monitoring. Thus, in this chapter, we are also concerned with the ability of BCUSUM to detect small shifts in process mean.

25.2.1 A likelihood ratio scheme

Gardner[19] noted that although the BCUSUM is considered to be the most thorough forecast monitoring procedure available, it has not been widely used in practice because of the lack of published control limits and studies analyzing its performance in comparison with the other established techniques. In this section, we derive the BCUSUM scheme using the likelihood ratio principle. Using this approach, we can easily determine the appropriate form of the control limits.

We assume that errors arising from a given forecasting model are independent and normally distributed with mean $\mu_\varepsilon = 0$ and constant variance σ_ε^2. Note that with this assumption, we can easily view the forecast errors as the data analogous to the process measurements $\{y_t\}$ used in SPC which are often assumed i.i.d. normal when no external sources of variation are affecting the process. In what follows, we discuss the BCUSUM scheme in the context of SPC measurements $\{y_t\}$. Without loss of generality, we assume that when the process is in control, $\{y_t\}$ has mean $\mu = 0$ and constant variance σ^2. Thus, in monitoring $\{y_t\}$, our interest is in sequentially testing the hypothesis

$$H_0 : \mu = 0, \quad \text{for} \quad y_1, y_2, \ldots, y_n, \tag{25.2}$$

versus the alternative hypothesis

$$H_j : \begin{cases} \mu = 0, & \text{for } y_1, y_2, \ldots, y_{j-1}, \\ \mu = \mu_1 (\mu_1 > 0), & \text{for } y_j, y_{j+1}, \ldots, y_n. \end{cases} \tag{25.3}$$

The likelihood ratio under the hypotheses H_0 and H_j is

$$\begin{aligned}\lambda_n(j) &= \frac{\Pi_{i=1}^{n} f_{\mu=0}(y_i)}{\Pi_{i=1}^{j-1} f_{\mu=0}(y_i)\Pi_{i=j}^{n} f_{\mu=\mu_1}(y_i)} \\ &= \frac{\Pi_{i=j}^{n} f_{\mu=0}(y_i)}{\Pi_{i=j}^{n} f_{\mu=\mu_1}(y_i)} \\ &= \frac{\Pi_{i=j}^{n}(2\pi)^{-1/2}\sigma^{-1}\exp\left[y_i^2/2\sigma^2\right]}{\Pi_{i=j}^{n}(2\pi)^{-1/2}\sigma^{-1}\exp\left[(y_i-\mu_1)^2/2\sigma^2\right]} \\ &= \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=j}^{n} y_i^2 - \sum_{i=j}^{n}(y_i-\mu_1)^2\right]\right\}.\end{aligned} \tag{25.4}$$

The rejection region, using the Neyman–Pearson lemma, is given by $\lambda_n(j) < k$. Taking the natural logarithm and simplifying (25.4), we therefore have the following rejection region for testing the null hypothesis (25.2) against the alternative hypothesis (25.3):

$$\frac{\sum_{i=j}^{n} y_i}{n-j+1} > \frac{-\sigma^2 \ln k}{\mu_1(n-j+1)} + \frac{\mu_1}{2}. \tag{25.5}$$

The left-hand side of this inequality is simply the average of the latest $n-j+1$ observations. We denote this average as $\bar{y}_j$. Under H_0, $\bar{y}_j$ is normally distributed with mean zero and variance $\sigma^2/(n-j+1)$. Since the quantity on the right-hand side of (25.5) is a constant (say, k'), it follows that $\bar{y}_j$ is the test statistic that yields the most powerful test and the rejection region is $\bar{y}_j > k'$. That is, the rejection region, RR, is given by

$$RR = \left\{\bar{y}_j > k'\right\}.$$

The precise value of k' is determined by fixing α and noting that

$$\alpha = P\left(\bar{y}_j \text{ in } RR \text{ when } \mu = \mu_0 = 0\right).$$

Note that the form taken by the rejection region does not depend upon the particular value assigned to μ_1. That is, any value of μ greater than $\mu_0 = 0$ would lead to exactly the same rejection region. This gives rise to a UMP test for testing (25.2) against

$$H_j = \begin{cases} \mu = 0, & \text{for} \quad y_1, y_2, \ldots, y_{j-1}, \\ \mu > \mu_1, & \text{for} \quad y_j, y_{j+1}, \ldots, y_n. \end{cases}$$

The test is just a simple z-test with rejection region

$$\frac{\sum_{i=j}^{n} y_i}{\sigma\sqrt{n-j+1}} \geq z_\alpha. \tag{25.6}$$

It is well known that the value of k' can be determined by fixing

$$\alpha = P\left[\frac{\sum_{i=j}^{n} y_i}{n-j+1} > k' \,|\mu = 0\right].$$

In SPC applications, the time j when the change occurred is unknown a priori. Thus, in order to detect whether there has been a change in the process average on an on-line basis, we need calculate the following backward standardized average (BCuZ):

$$\text{BCuZ}_j = \frac{\sum_{i=j}^{n} y_i}{\sigma\sqrt{n-j+1}}, \tag{25.7}$$

for $1 \le j \le n$. When a particular BCuZ exceeds a pre-specified threshold value, say z^*, we conclude that there has been a change in the process average. Note that at every time period t, we calculate n dependent BCuZs. Thus, unlike the z_α in (25.6), z^* is the threshold for the n dependent BCuZs calculated at time t. For the BCuZ, the α-value in (25.6) cannot therefore be interpreted as the probability of eventually obtaining a false alarm. One chooses z^* such that the control charting scheme produces a desired in-control ARL. This can be done using simulation.

Instead of plotting the BCuZ, one may utilize the BCUSUM with the corresponding control limits

$$\text{BCUSUM}_j = \sum_{i=j}^{n} y_i \ge z^*\sigma\sqrt{n-j+1}, \quad 1 \le j \le n. \tag{25.8}$$

One can see that the control limits of the above BCUSUM resemble a parabolic mask. This BCUSUM can also be implemented using a backward moving average (BMA) scheme. The BMA with its control limits takes the form

$$\text{BMA}_j = \frac{\sum_{i=j}^{n} y_i}{n-j+1} \ge z^* \frac{\sigma}{\sqrt{n-j+1}}, \quad 1 \le j \le n.$$

In using the BCuZ, BCUSUM or BMA, one needs to plot n points at each time period t. If one prefers a Shewhart-like plot, the following statistic may be monitored at each time period:

$$\max_{1\le j\le n}\left[\frac{\sum_{i=j}^{n} y_i}{\sigma\sqrt{n-j+1}}\right].$$

When this statistic exceeds z^*, one concludes that the process average has shifted. We can see that the above statistic is a special case of the level shift detection procedure described by Tsay.[22]

For a two-sided alternative for hypothesis (25.2) where j is known, we use the following statistic to determine whether there has been a change in mean:

$$\left|\frac{\sum_{i=j}^{n} y_i}{\sigma\sqrt{n-j+1}}\right| \ge z_{\alpha/2}.$$

For on-line detection of mean shift, we may use the following BCUSUM scheme:

$$\begin{aligned} \text{BCUSUM}_j &= \sum_{i=j}^{n} y_i \ge z^*\sigma\sqrt{n-j+1} \\ \text{BCUSUM}_j &= \sum_{i=j}^{n} y_i \le -z^*\sigma\sqrt{n-j+1} \end{aligned}, \; 1 \le j \le n, \tag{25.9}$$

in which the right-hand side represents a parabolic mask. Alternatively, one may monitor the following statistic:

$$\max_{1\le j\le n}\left|\frac{\sum_{i=j}^{n} y_i}{\sigma\sqrt{n-j+1}}\right| \ge z^*. \tag{25.10}$$

Similar to the one-sided monitoring scheme, the z^* for equations (25.9) and (25.10) is chosen such that the in-control ARL is equivalent to some pre-specified value.

25.3 SYMMETRIC CUMULATIVE SUM SCHEMES

The BCUSUM scheme discussed earlier is computationally intensive. Interestingly, we can translate the BCUSUM scheme into an equivalent CUSUM representation. The following characteristics of symmetric functions are important in establishing a CUSUM scheme based on the UMP test. Let $y = f_1(t)$ and $y = f_2(t)$ represent two real functions of t. If $f_1(t)$ and $f_2(t)$ are symmetric about $y = 0$, then $f_1(t) = -f_2(t)$. In general, when $f_1(t)$ and $f_2(t)$ are symmetric about $y = a$, then

$$\begin{aligned} f_1(t) &= f_2(t) - 2[f_2(t) - a] \\ &= 2a - f_2(t). \end{aligned} \tag{25.11}$$

The relationship given by equation (25.11) is shown in Figure 25.1.

Let $j = n$, where n is an integer greater than or equal to 1. Then, for $1 \le j \le n$, the cumulative sum can be written as

$$\begin{aligned} \text{CUSUM}_j &= \sum_{i=1}^{j} y_i = \sum_{i=1}^{n} y_i - \sum_{i=j+1}^{n} y_i, \\ &= \text{CUSUM}_n - \sum_{i=j+1}^{n} y_i \end{aligned} \tag{25.12}$$

The term $\sum_{i=j+1}^{n} y_i$ calculates the cumulative sum in a backward fashion. Hence, we call this term the backward CUSUM or BCUSUM. Thus, (25.12) can be written as

$$\text{CUSUM}_j = \text{CUSUM}_n - \text{BCUSUM}_{j+1}, \quad 1 \le j \le n. \tag{25.13}$$

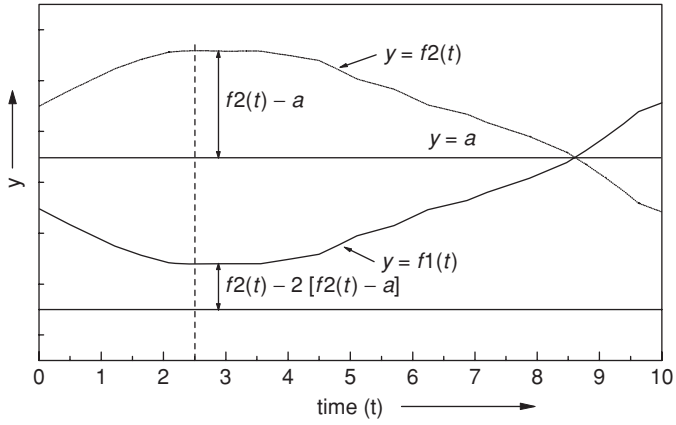


Figure 25.1 A plot of symmetric functions $f_1(t)$ and $f_2(t)$.

Note that in the above equation, $\text{BCUSUM}_{j+1} = 0$ for $j+1 > n$. From (25.13) it is evident that CUSUM_j and BCUSUM_{j+1} are symmetric and the point of symmetry is at $\text{CUSUM}_n/2$. By the virtue of their symmetry, one can therefore implement the CUSUM mask for the BCUSUM scheme and vice versa. In the following section, we will derive a BCUSUM mask based on a UMP test.

25.3.1 Parabolic mask for CUSUM

The BCUSUM limit for the $(j+1)$th observation is $z^*\sqrt{n-j}$. By symmetry, we can therefore derive the following CUSUM scheme for detecting an increase in the mean:

$$\text{CUSUM}_j \leq -z^*\sigma\sqrt{n-j} + \text{CUSUM}_n. \tag{25.14}$$

For a two-side scheme one may implement:

$$\begin{aligned} \text{CUSUM}_j &\geq z^*\,\sigma\sqrt{n-j} + \text{CUSUM}_n, \\ \text{CUSUM}_j &\leq -z^*\,\sigma\sqrt{n-j} + \text{CUSUM}_n. \end{aligned} \tag{25.15}$$

Similar to the BCUSUM scheme, the value of z^* in (25.14) is chosen such that the CUSUM achieves a pre-specified in-control ARL.

In using (25.14), we assume that $\text{CUSUM}_0 = 0$. With this assumption, the origin is also used to decide an out-of-control situation (i.e. it must also lie within the CUSUM mask when the process is in control). For a V-mask, for example, if the first observation is greater than $h+k$, the origin is the point that will generate the signal. Similarly, for the parabolic mask of (25.14), the first observation will signal a change when the y_1 is greater than z^*.

25.3.2 Numerical illustration

To throw more light on the relationship between the CUSUM and BCUSUM, we illustrate the results in the preceding section using a numerical example. The data in Table 25.2 were obtained from Woodall and Adams.[23] In the table, the first 10 observations are random variables from $N(0,1)$. The next five observations are from $N(1,1)$. The CUSUM is calculated using the conventional method while BCUSUM values are calculated using equation (25.8); the corresponding masks are computed using equations (25.4) and (25.8), respectively.

Visually analyzing the data in Table 25.2, one cannot easily detect the symmetry of the CUSUM and BCUSUM. Their relationship becomes obvious once plotted against time (see Figure 25.2). From Figure 25.2, one can easily appreciate the symmetry of the two CUSUM schemes which is used to establish the parabolic mask for the CUSUM.

25.3.3 ARL performance of CUSUM when the BCUSUM mask is used

Atienza *et al.*[15] compare the ARL performance of a CUSUM parabolic mask with other CUSUM mask schemes. By using Monte Carlo simulation with 5000 iterations, they give the ARL and standard deviation of ARL (SARL) of the various CUSUM schemes for selected σ shifts in mean. The sequence of normal random variables was generated using the IMSL[24] Statistical Library. The comparisons are replicated in Table 25.3.

Table 25.2 Calculation of CUSUM and BCUSUM with the corresponding masks.

Time (j)	Observation y_j	BCUSUM$_j$	CUSUM$_j$	BCUSUM mask	CUSUM mask
1	0.54	10.08	0.54	12.55	−2.04
2	−1.20	9.54	−0.66	12.12	−1.60
3	2.20	10.74	1.54	11.68	−1.14
4	−1.24	8.54	0.30	11.22	−0.67
5	−0.56	9.78	−0.26	10.75	−0.17
6	1.30	10.34	1.04	10.25	0.36
7	−0.02	9.04	1.02	9.72	0.92
8	0.14	9.06	1.16	9.16	1.51
9	0.16	8.92	1.32	8.57	2.14
10	0.86	8.76	2.18	7.94	2.84
11	0.92	7.90	3.10	7.24	3.60
12	2.40	6.98	5.50	6.48	4.47
13	3.08	4.58	8.58	5.61	5.50
14	0.26	1.50	8.84	4.58	6.84
15	1.24	1.24	10.08	3.24	10.08

For comparison purposes, the control chart parameters are chosen such that the in-control ARL for the one-sided scheme is approximately 470. As expected, the BCUSUM and CUSUM with parabolic mask give similar ARL performance. From the table, it is clear that for a shift in mean below $0.5\,\sigma$ and greater than 1.5σ, the CUSUM with parabolic mask gives the best ARL performance. The CUSUM with V-mask is superior only in tracking down shifts it is designed to detect (i.e., $2k$). The Shewhart CUSUM scheme[25] is an improvement on the sensitivity of the equivalent V-mask scheme but it is still less sensitive compared to a CUSUM with a parabolic, Bissell or snub-nosed mask.

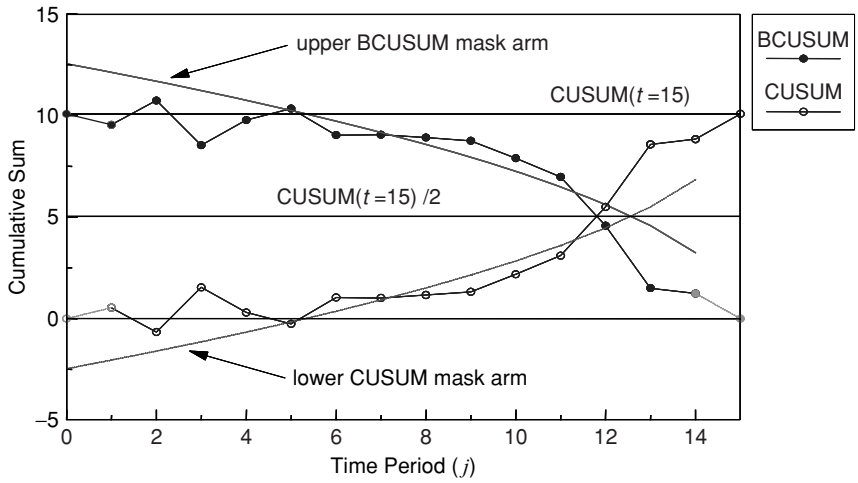


Figure 25.2 Plot of CUSUM$_j$ and BCUSUM$_{j+1}$ using Woodall and Adams'[23] data. Note that the CUSUM is plotted against j while the BCUSUM is plotted against $j+1$.

Table 25.3 ARL and SARL comparisons for different one-sided process monitoring schemes (from table 2, Atienza *at al.*[15]).

	V-mask		S-CUSUM		Bissell's mask		Snub-nosed mask		BCUSUM		Parabolic mask	
Shift σ	ARL	(SARL)	ARL	(SARL)	ARL	(SARL)	ARL	(SARL)	ARL	(SARL)	ARL	(SARL)
0.00	476.63	(465.68)	471.11	(458.15)	470.76	(468.50)	471.19	(473.14)	474.17	(467.05)	478.64	(478.49)
0.25	92.82	(88.46)	92.52	(88.31)	113.63	(106.57)	104.36	(98.81)	86.46	(67.95)	83.74	(65.18)
0.50	29.95	(24.46)	29.90	(24.42)	33.79	(27.65)	32.91	(27.99)	29.80	(20.20)	30.15	(20.00)
0.75	14.52	(9.42)	14.52	(9.42)	16.17	(11.01)	15.30	(10.32)	15.40	(9.52)	15.49	(9.52)
1.00	9.06	(4.85)	9.03	(4.85)	9.59	(5.50)	9.34	(5.41)	9.64	(5.58)	9.79	(5.61)
1.25	6.48	(3.05)	6.46	(3.03)	6.78	(3.54)	6.65	(3.29)	6.69	(3.61)	6.73	(3.64)
1.50	5.07	(2.11)	5.06	(2.12)	5.09	(2.58)	5.09	(2.44)	5.07	(2.61)	5.00	(2.58)
1.75	4.17	(1.54)	4.14	(1.57)	3.99	(2.01)	3.97	(1.86)	3.94	(1.95)	3.94	(1.97)
2.00	3.55	(1.21)	3.50	(1.24)	3.27	(1.61)	3.38	(1.56)	3.17	(1.50)	3.17	(1.51)
2.25	3.09	(0.97)	3.04	(1.06)	2.70	(1.35)	2.77	(1.28)	2.65	(1.23)	2.68	(1.25)
2.50	2.78	(0.83)	2.71	(0.93)	2.30	(1.12)	2.38	(1.09)	2.26	(1.03)	2.29	(1.03)
2.75	2.54	(0.69)	2.43	(0.84)	1.95	(0.92)	2.07	(0.91)	1.98	(0.87)	1.98	(0.86)
3.00	2.32	(0.60)	2.16	(0.77)	1.73	(0.80)	1.81	(0.80)	1.77	(0.77)	1.78	(0.79)
3.25	2.16	(0.53)	1.96	(0.68)	1.53	(0.67)	1.62	(0.69)	1.58	(0.67)	1.57	(0.65)
3.50	2.02	(0.46)	1.77	(0.64)	1.39	(0.58)	1.46	(0.60)	1.44	(0.58)	1.44	(0.58)
3.75	1.92	(0.44)	1.64	(0.60)	1.29	(0.50)	1.35	(0.53)	1.31	(0.50)	1.33	(0.51)
4.00	1.83	(0.43)	1.48	(0.55)	1.20	(0.41)	1.24	(0.44)	1.24	(0.45)	1.24	(0.44)

Notes: V-mask, $h = 4.33$, $k = 0.5$; Shewhart CUSUM, $h = 4.33$, $k = 0.5$; Shewhart UCL= +3.90; snub-nosed V-Mask, $h_1 = 1.89$, $k_1 = 1.40$, $h_2 = 4.66$, $k_2 = 0.50$; BCUSUM, $z^* = 3.24$; parabolic mask, $z^* = 3.24$.

25.4 CUSUM SCHEME FOR AUTOCORRELATED OBSERVATIONS

The issues associated with monitoring autocorrelated processes have been presented in Chapters 22 and 23, and some of them which have been addressed further in this and the previous chapter. Here, using the characteristics of BCUSUM, a CUSUM scheme for detecting a step change in process mean that directly utilizes the autocorrelated process observations is presented.

25.4.1 Formulation

Assuming that the process is stable, $\{y_t\}$ can be described by a stationary ARMA process as given in equation (25.1). In SPC, our objective is to detect changes in mean as early as possible. Recall that the sum or average of all the measurements since the change occurred is the best indicator of process change, if we know or can guess the time of occurrence. This intuitive claim can be supported by the hypothesis testing problem shown in (25.2) and (25.3).

In the presence of autocorrelation, under H_0 in (25.2), we have

$$\bar{y}_{n-j+1}\sqrt{n-j+1} \sim N\left(0, \sum_{|h|<n-j+1}\left(1-\frac{|h|}{n-j+1}\right)\gamma(h)\right), \tag{25.16}$$

where

$$\bar{y}_{n-j+1} = \frac{\sum_{i=j}^{n} y_i}{n-j+1}$$

and $\gamma(\cdot)$ is the autocovariance function of $\{y_t,\ t=1,\ 2,\ \ldots\}$.[26] The variance of $\bar{y}_{n-j+1}\sqrt{n-j+1}$ approaches $\upsilon^2 = \sum_{h=-\infty}^{\infty}\gamma(h)$ as n becomes large. Using (25.16), we can therefore test (25.2) against (25.3) using the test statistic

$$z = \frac{\bar{y}_{n-j+1}\sqrt{n-j+1}}{\sqrt{\sum_{|h|<n-j+1}(1-|h|/n-j+1)\,\gamma(h)}}. \tag{25.17}$$

With the inclusion of the autocorrelation coefficient, the development is similar to that of the i.i.d. case. We reject the null hypothesis (25.2) in favor of (25.3) when the z-statistic in (25.17) exceeds a certain critical value z_α, that is,

$$\frac{\bar{y}\sqrt{n-j+1}}{\sqrt{\sum_{|h|<n-j+1}(1-|h|/n-j+1)\,\gamma(h)}} > z_\alpha. \tag{25.18}$$

As the exact time of occurrence of change is not known a priori, to implement the scheme on an on-line basis, we need to calculate n z-statistics,

$$z_j = \frac{\bar{y}_{n-j+1}\sqrt{n-j+1}}{\sqrt{\sum_{|h|<n-j+1}(1-|h|/n-j+1)\,\gamma(h)}}, \qquad j = 1, 2, \ldots, n. \tag{25.19}$$

Similarly, when a z_j exceeds a pre-specified control limit, say z^* (to be obtained from simulation with a desired ARL value) we conclude that there has been a change in

mean; and for the same reason as in i.i.d. case, z_α in (25.18) is not used as the α-value cannot be interpreted as the probability of eventually obtaining a false alarm.[27]

Note that (25.19) can also be written as

$$\sum_{i=j}^{n} y_i > z_\alpha \sqrt{(n-j+1)\sum_{|h|<n-j+1}\left(1-\frac{|h|}{n-j+1}\right)\gamma(h)}, \qquad j=1,2,\ldots,n.$$

The left-hand side is just the sum of the $n-j+1$ most recent observations, which is the (BCUSUM) shown in the previous section. We can therefore test the presence of a step shift in mean by applying the following scheme:

$$\text{BCUSUM}_j^n \geq z^* \upsilon_{n-j+1}\sqrt{n-j+1}, \qquad j=1,2,\ldots,n, \tag{25.20}$$

where

$$\upsilon_{n-j+1}^2 = \sum_{|h|<n-j+1}\left(1-\frac{|h|}{n-j+1}\right)\gamma(h).$$

Similar to (25.9), to detect both an increase and a decrease in mean, the following scheme may be implemented:

$$\begin{aligned} \text{BCUSUM}_j^n &\geq z^* \upsilon_{n-j+1}\sqrt{n-j+1}, \\ \text{BCUSUM}_j^n &\leq -z^* \upsilon_{n-j+1}\sqrt{n-j+1}, \end{aligned} \quad j=1,2,\ldots,n. \tag{25.21}$$

Graphically, the control limits for the BCUSUM in (25.20) and (25.21) resemble a parabolic mask. Note that in implementing the above BCUSUM schemes, one only needs an estimate of the autocovariance function of the series being monitored.

25.4.2 Mask representation

Implementing the BCUSUM from either equation (25.20) or (25.21) for the on-line detection of mean shift implies that we need to calculate n BCUSUMs at every time period t. In addition, note that n increases as more observations become available. The BCUSUM procedure is therefore computationally intensive when implemented on-line. Interestingly, we can express the BCUSUM decision rules shown in equations (25.20) and (25.21) using the typical CUSUM representation, which has been shown Section 25.3 (see the derivations of equation (25.13)) to be as follows:

$$\text{CUSUM}_j^n \leq -z^* \upsilon_{n-j}\sqrt{n-j} + \text{CUSUM}_n^n, \qquad j=1,2,\ldots,n, \tag{25.22}$$

and

$$\begin{aligned} \text{CUSUM}_j^n &\geq z^* \upsilon_{n-j}\sqrt{n-j} + \text{CUSUM}_n^n, \\ \text{CUSUM}_j^n &\leq -z^* \upsilon_{n-j}\sqrt{n-j} + \text{CUSUM}_n^n, \end{aligned} \quad j=1,2,\ldots,n, \tag{25.23}$$

respectively. Similar to the BCUSUM scheme, the z^* in equation (25.22) or (25.23) is chosen such that the CUSUM chart achieves a pre-specified in-control ARL. The parabolic CUSUM scheme for monitoring a sequence of i.i.d. normal random variables shown in Section 25.3 is a special case of equation (25.23).

In using equation (25.22) or (25.23), we also assume that $\text{CUSUM}_0^0 = 0$ as in the i.i.d. case. Similarly, for the parabolic masks equation (25.22) or (25.23), the first observation

will signal a change when $y_1 > z^* \upsilon_1$. The assumption that $\text{CUSUM}_0^0 = 0$ is crucial to the use of the V-mask in which the initial observation is used in decision making applies only when initial CUSUM value is zero and for a stationary process with a zero mean. When the process is not centered at zero, one monitors the following the following CUSUM scheme:

$$\text{CUSUM}_j^n = \sum_{i=1}^{j} (y_i - \mu_0),$$

where μ_0 represents the mean of the process.

25.4.3 Sensitivity analysis

In analyzing the performance of the proposed CUSUM scheme, we focus our attention on the detection of changes in the mean of an AR(1) process. The importance of an AR(1) process in SPC has been emphasized in the literature.[8,11,12] In the following, we describe a simple graphical procedure that can facilitate the choice of the constraint z^*. Subsequently, we compare the performance of the proposed CUSUM scheme with the other established procedures for monitoring autocorrelated processes. We then address one of the major criticisms in implementing a CUSUM in mask form (i.e. how far should we extend the CUSUM mask arms?).

25.4.3.1 Choice of z^*

Determining the CUSUM parameter z^* through simulation may take a lot of time specially when the initial guess for z^* is far from the required value. One simple approach is to construct a monogram similar to that shown in Figure 25.3.

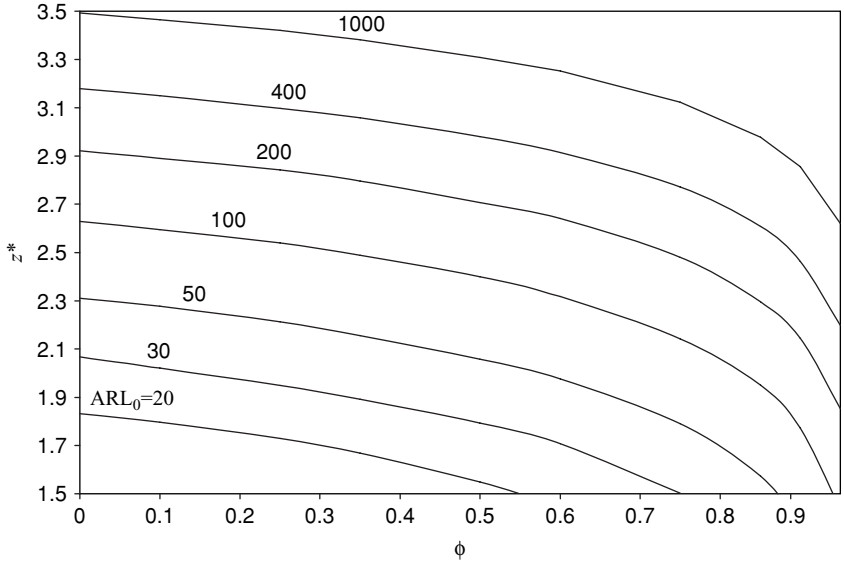


Figure 25.3 Choice of z^* for various combinations of ARL and ϕ.

Let δ_ϕ represent the error in estimating the AR(1) parameter ϕ. Using Figure 25.3 and given a target in-control ARL, one can see that the effect of δ_ϕ on the appropriate z^* is less significant for ϕ near 0 compared to when ϕ is near 1. This signifies that misspecification of ϕ would not result in a significantly different z^* when the autocorrelation is weak. However, for a highly positively autocorrelated process, small misspecification of ϕ would result in a significantly different z^*.

Since the parabolic mask that dictates the properties of the proposed CUSUM scheme is characterized not only by z^* but also by the autocovariance structure of the assumed model, it would be interesting to see the different parabolic masks for various values of ϕ that would produce a given in-control ARL. The parabolic masks that would provide an in-control ARL of 370 for a two-sided CUSUM scheme (or $ARL_0 = 740$ for a one-sided CUSUM scheme) various values of ϕ are shown in Figure 25.4. From Figure 25.4, one can see that as ϕ approaches 1 the change in the angle of the arms of the parabola becomes faster. Thus, one can conclude that the proposed CUSUM scheme is less sensitive to misspecification of ϕ when the process is weakly positively autocorrelated compared to when the process is highly positively autocorrelated.

Note that when ϕ is overestimated, the resulting parabola has a wider envelope than expected. The overestimation of ϕ is therefore expected to produce a CUSUM scheme

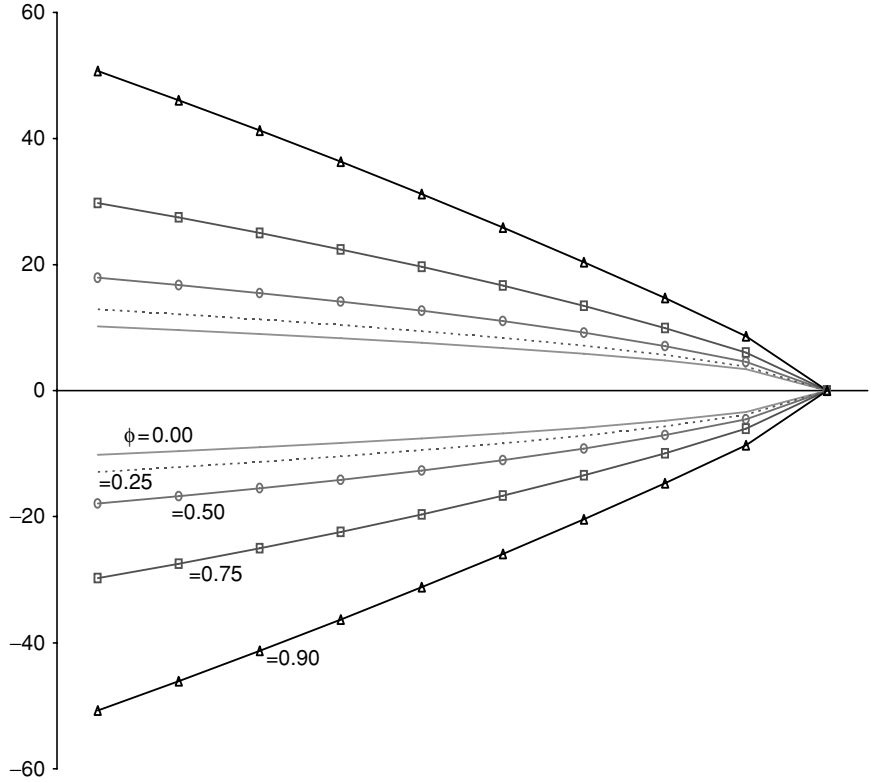


Figure 25.4 Parabolic masks for selected AR processes ($\phi = 0.00, 0.25, 0.50, 0.75$ and 0.90); in-control ARL = 370.

that has a much larger ARL than expected. Conversely, when ϕ is underestimated, the resulting parabola has a narrower envelope than expected. This will result in a smaller than expected ARL. This property of the proposed CUSUM is similar to the properties of the CUSUM or exponentially weighted moving average applied to forecast residuals[27]).

25.4.3.2 ARL comparisons

For various levels of autocorrelation ϕ_1 and shift in mean δ, Atienza *et al*.[28] compare the ARL performance of the proposed CUSUM with the SCR and CUSUMR. Their results are replicated in Table 25.4. The shift in mean is measured in terms of both σ_ε and σ_y. For comparison purposes, the control chart parameters for SCR, CUSUMR and the proposed CUSUM are chosen such that the in-control ARL for the one-sided scheme is approximately 740.

From Table 25.4, it can be seen that the performance of the proposed CUSUM is very close to the performance of the CUSUM on residuals. For large process shifts, the proposed CUSUM outperforms both the SCR and CUSUMR. In detecting small shifts for highly positively autocorrelated processes, CUSUMR is more sensitive than the proposed CUSUM. Note, however, that for ϕ_1 close to 1 (i.e. highly positively autocorrelated processes) a process shift measured in term of σ_ε is small compared to when the shift is measured in terms of σ_y. For example, a $2\sigma_\varepsilon$ shift for an AR(1) process with $\phi_1 = 0.75$ is equal to only a 1.323 σ_y shift. Similarly, when $\sigma_1 = 0.9$, a $2\sigma_\varepsilon$ shift in mean is equivalent to only a 0.872 σ_y shift. By this token, we can expect that the proposed CUSUM will outperform CUSUMR in a wider shift range when the shift is measured in terms of σ_y.

25.4.3.3 Length of the mask arm

Ideally, in implementing the CUSUM in a mask scheme, one needs to extend the mask arms to cover all the data points. Thus, when the process is stable, we will often need to store a large amount data for process monitoring. To determine the effect of the length of mask arms on the sensitivity of the proposed CUSUM in detecting a shift in mean, Atienza *et al*.[28] analyze the performance of the CUSUM when the analysis is based only on the latest *mw* (moving window) of process observations. For a particular *mw* size, the value of z^* is adjusted such that the in-control ARL will be approximately 740. The results are taken from Atienza *et al*. [28] and shown in Table 25.5. Note that the ARL figures in the table are calculated using the same approach as described in the previous section.

From Table 25.5, we can see that for small values of ϕ_1, we may focus our attention on the latest 50 observations without severely compromising the sensitivity of the CUSUM scheme. For larger values of ϕ_1, we need to concentrate on the latest 100 observations. For all cases, the performance of the CUSUM scheme utilizing only the latest 200 observations is already quite close to the performance of the CUSUM that uses all process observations. From these results, it is quite clear that one does not need to indefinitely extend CUSUM mask arm in order to effectively detect shifts in process mean.

Table 25.4 ARL (SARL) Comparison for one-sided SCR, CUSUMR and CUSUM schemes (from Table 1 of Atienza et al.[28]).

			δ (in σ_ε)							
Φ_1	Control Chart	Chart Parameter(s)	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00
0.00		δ(in σ_x)	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00
	SCR	UCL= +3.00	740.78	335.60	161.04	81.80	43.96	14.97	6.30	2.00
			(740.26)	(335.09)	(160.54)	(81.30)	(43.45)	(14.46)	(5.78)	(1.41)
	CUSUMR	$h = 4.78, k = 0.50$	740.03	123.98	35.29	16.19	9.93	5.52	3.86	2.49
			(734.93)	(118.74)	(29.00)	(10.67)	(5.30)	(2.22)	(1.25)	(0.64)
	CUSUM	$z^* = 3.391$	740.92	102.45	34.47	17.27	10.69	5.44	3.46	1.88
			(737.41)	(77.14)	(22.21)	(10.33)	(5.94)	(2.73)	(1.59)	(0.79)
0.25		δ(in σ_x)	0.00	0.24	0.48	0.73	0.97	1.45	1.94	2.9
	SCR	UCL= +3.00	740.78	406.64	230.38	134.58	80.94	31.70	13.59	3.21
			(740.26)	(406.34)	(230.31)	(134.73)	(81.29)	(32.33)	(14.29)	(3.42)
	CUSUMR	$h = 6.00, k = 0.375$	739.96	156.04	51.13	24.72	15.29	8.35	5.67	3.43
			(730.15)	(148.43)	(41.64)	(16.45)	(8.47)	(3.63)	(2.05)	(0.98)
	CUSUM	$z^* = 3.316$	740.74	151.57	53.65	27.13	16.52	8.07	4.71	2.22
			(741.44)	(120.94)	(37.66)	(17.87)	(10.12)	(4.78)	(2.73)	(1.21)

0.50		δ (in σ_x)	0.00	0.22	0.43	0.65	0.87	1.30	1.73	2.60
	SCR	UCL= +3.00	740.78	494.54	334.51	229.00	158.38	77.34	37.98	8.48
			(740.26)	(494.50)	(335.09)	(230.30)	(160.49)	(81.04)	(42.62)	(11.81)
	CUSUMR	$h = 8.02, k = 0.25$	740.02	211.55	83.83	44.09	28.09	15.44	10.33	5.98
			(728.30)	(193.50)	(69.66)	(31.34)	(16.77)	(7.22)	(4.13)	(1.90)
	CUSUM	$z^* = 3.205$	741.20	235.22	96.49	50.41	31.73	14.69	8.55	3.43
			(736.83)	(200.84)	(73.08)	(36.20)	(21.68)	(10.13)	(5.94)	(2.48)
0.75		δ(in σ_x)	0.00	0.17	0.33	0.50	0.66	0.99	1.32	1.98
	SCR	UCL= +3.00	740.78	603.67	492.95	402.88	328.96	216.40	136.49	41.90
			(740.26)	(603.94)	(494.49)	(406.30)	(334.95)	(229.41)	(156.97)	(64.67)
	CUSUMR	$h = 12.11, k = 0.125$	739.97	330.54	175.15	107.63	73.98	43.13	29.36	16.95
			(710.16)	(300.28)	(145.46)	(79.15)	(48.70)	(23.14)	(13.56)	(6.56)
	CUSUM	$z^* = 3.017$	741.57	394.21	217.06	127.79	87.03	44.20	25.54	10.33
			(744.00)	(370.38)	(192.81)	(105.02)	(70.65)	(34.51)	(20.56)	(9.28)
0.90		δ (in σ_x)	0.00	0.11	0.22	0.33	0.44	0.65	0.87	1.31
	SCR	UCL= +3.00	740.78	681.58	626.47	574.53	524.77	427.91	330.28	145.22
			(740.26)	(682.09)	(628.84)	(580.03)	(535.20)	(455.07)	(381.89)	(228.02)
	CUSUMR	$h = 17.75, k = 0.05$	739.99	499.22	355.22	265.14	206.18	137.13	99.66	61.27
			(691.52)	(437.77)	(297.32)	(206.46)	(146.85)	(86.80)	(57.11)	(28.91)
	CUSUM	$z^* = 2.734$	740.53	542.05	409.33	318.23	230.90	142.67	97.90	46.37
			(736.47)	(533.93)	(405.81)	(308.90)	(215.76)	(132.31)	(90.66)	(45.47)

Table 25.5 Effect of mw on the ARL(SRL) performance of the proposed CUSUM scheme (from Table 2 Atienza et al.[28]).

			Shift in mean (in σ_ε)								
ϕ_1	mw	z*	0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00
0.00	10	3.342	740.92	179.59	52.14	20.64	11.35	5.32	3.35	2.38	1.84
			(740.34)	(175.14)	(48.26)	(17.06)	(7.96)	(2.75)	(1.57)	(1.05)	(0.78)
	25	3.372	741.20	140.77	38.93	17.28	10.40	5.39	3.39	2.40	1.86
			(743.31)	(135.06)	(31.38)	(11.54)	(5.98)	(2.68)	(1.58)	(1.07)	(0.79)
	50	3.383	740.21	121.77	35.04	17.26	10.58	5.48	3.42	2.43	1.88
			(744.49)	(110.26)	(24.64)	(10.46)	(6.02)	(2.73)	(1.61)	(1.08)	(0.78)
	100	3.388	740.14	109.56	34.28	17.22	10.62	5.47	3.43	2.44	1.88
			(742.00)	(93.19)	(22.31)	(10.41)	(5.97)	(2.73)	(1.61)	(1.09)	(0.78)
	200	3.390	740.92	103.73	34.43	17.27	10.68	5.44	3.42	2.42	1.88
			(739.60)	(81.23)	(22.16)	(10.37)	(5.94)	(2.73)	(1.59)	(1.08)	(0.79)
	All	3.391	740.92	102.45	34.47	17.27	10.69	5.44	3.46	2.43	1.88
			(737.41)	(77.14)	(22.21)	(10.33)	(5.94)	(2.73)	(1.59)	(1.08)	(0.79)
0.25	10	3.261	740.96	238.83	88.18	38.17	19.88	7.95	4.64	2.98	2.14
			(741.18)	(231.27)	(85.46)	(35.73)	(16.66)	(5.29)	(2.77)	(1.72)	(1.20)
	25	3.293	740.21	204.99	66.43	29.50	16.69	8.04	4.72	3.04	2.18
			(747.92)	(197.08)	(59.27)	(23.18)	(11.42)	(4.80)	(2.73)	(1.71)	(1.20)
	50	3.306	741.24	183.20	57.88	27.04	16.23	8.02	4.68	3.10	2.21
			(744.47)	(171.59)	(48.04)	(18.51)	(10.12)	(4.78)	(2.73)	(1.76)	(1.22)
	100	3.312	740.99	166.07	54.23	26.85	16.54	8.02	4.69	3.08	2.22
			(743.18)	(149.69)	(40.46)	(17.69)	(10.19)	(4.75)	(2.75)	(1.76)	(1.22)
	200	3.314	740.21	155.38	53.43	27.11	16.46	8.06	4.72	3.09	2.22
			(739.16)	(132.76)	(37.46)	(17.91)	(10.08)	(4.75)	(2.74)	(1.78)	(1.24)
	All	3.316	740.74	151.57	53.65	27.13	16.52	8.07	4.71	3.12	2.22
			(741.44)	(120.94)	(37.66)	(17.87)	(10.12)	(4.78)	(2.73)	(1.78)	(1.21)

			Shift in Mean (in σ_ε)								
ϕ_1	mw	z^*	0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00
0.50	10	3.153	740.92	322.19	150.31	78.83	43.90	16.81	8.69	5.12	3.28
			(749.28)	(310.89)	(151.31)	(75.15)	(41.57)	(15.09)	(7.27)	(4.02)	(2.45)
	25	3.182	741.24	296.54	125.80	63.79	35.12	14.74	8.28	5.13	3.34
			(747.78)	(287.54)	(123.97)	(57.70)	(30.39)	(10.60)	(5.83)	(3.73)	(2.48)
	50	3.194	741.43	275.27	111.07	55.27	32.11	14.62	8.32	5.12	3.39
			(743.17)	(262.46)	(104.11)	(46.75)	(24.49)	(9.96)	(5.76)	(3.68)	(2.50)
	100	3.200	741.43	256.69	100.43	51.46	31.60	14.72	8.50	5.19	3.40
			(742.65)	(238.85)	(85.67)	(38.74)	(21.79)	(10.10)	(5.90)	(3.76)	(2.49)
	200	3.203	740.96	244.38	97.08	50.43	31.61	14.73	8.54	5.21	3.43
			(739.25)	(218.18)	(76.38)	(36.30)	(21.73)	(10.10)	(5.96)	(3.75)	(2.49)
	All	3.205	741.20	235.22	96.49	50.41	31.73	14.69	8.55	5.20	3.43
			(736.83)	(200.84)	(73.08)	(36.20)	(21.68)	(10.13)	(5.94)	(3.78)	(2.48)
0.75	10	2.969	740.21	440.64	274.63	185.16	122.61	58.62	30.32	17.94	11.04
			(735.20)	(434.81)	(275.77)	(184.30)	(120.52)	(59.31)	(31.19)	(17.85)	(11.57)
	25	2.991	740.34	433.78	260.48	168.44	111.28	51.77	27.22	15.98	10.33
			(744.89)	(433.13)	(262.38)	(167.34)	(108.08)	(41.63)	(26.28)	(14.75)	(10.01)
	50	3.004	740.14	419.40	248.51	151.96	100.90	47.00	25.55	15.92	10.21
			(743.44)	(410.78)	(247.73)	(148.20)	(95.85)	(41.43)	(21.91)	(13.66)	(9.39)
	100	3.012	740.21	408.27	233.90	138.99	92.20	44.70	25.35	16.09	10.21
			(740.45)	(398.56)	(229.58)	(125.88)	(83.24)	(36.43)	(20.59)	(13.74)	(9.29)
	200	3.015	741.24	403.04	223.93	131.21	86.93	44.19	25.40	16.18	10.34
			(745.10)	(388.40)	(211.27)	(112.79)	(72.07)	(34.57)	(20.58)	(13.81)	(9.29)
	All	3.017	741.57	394.21	217.06	127.79	87.03	44.20	25.54	16.24	10.33
			(744.00)	(370.38)	(192.81)	(105.02)	(70.65)	(34.51)	(20.56)	(13.73)	(9.28)

(*Continued*)

Table 25.5 Effect of mw on the ARL(SRL) performance of the proposed CUSUM scheme (from Table 2 Atienza et al.[28]). (*Contineued*)

			Shift in Mean (in σ_ε)								
ϕ_1	mw	z^*	0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00
0.90	10	2.709	740.53	555.85	426.97	338.67	254.69	162.43	114.02	71.72	51.33
			(749.76)	(553.89)	(435.00)	(345.92)	(257.55)	(173.15)	(117.10)	(76.82)	(55.73)
	25	2.716	740.49	557.07	425.04	337.30	254.26	160.80	113.38	71.06	49.88
			(755.12)	(553.57)	(429.05)	(348.23)	(256.05)	(168.92)	(116.09)	(75.86)	(53.27)
	50	2.723	740.53	554.90	424.90	332.30	250.08	156.41	109.43	68.83	48.26
			(747.07)	(549.57)	(430.99)	(342.59)	(249.30)	(161.72)	(113.50)	(72.70)	(50.64)
	100	2.729	740.96	551.67	421.02	329.14	243.27	151.39	104.42	66.46	45.98
			(739.88)	(546.87)	(425.45)	(337.90)	(240.56)	(152.94)	(103.74)	(66.43)	(46.28)
	200	2.732	740.65	547.06	414.75	323.74	237.52	145.41	100.59	65.28	45.52
			(738.82)	(540.14)	(414.57)	(326.70)	(231.39)	(143.04)	(97.07)	(63.36)	(44.92)
	All	2.734	740.53	542.05	409.33	318.23	230.90	142.67	97.90	65.98	46.37
			(736.47)	(533.93)	(405.81)	(308.90)	(215.76)	(132.31)	(90.66)	(63.16)	(45.47)

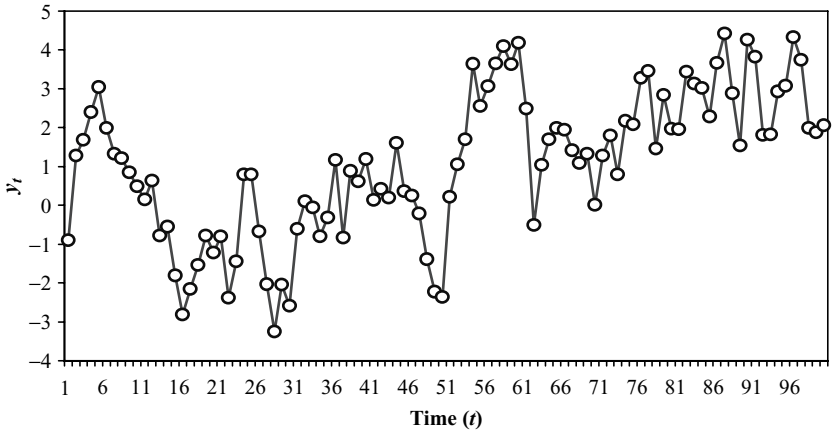


Figure 25.5 The simulated AR(1) series with $\phi_1 = 0.75$, $\mu = 0$, and $\sigma_\varepsilon = 1$.

25.4.4 An example

To illustrate the application of the proposed CUSUM scheme, Atienza *et al*.[28] simulated the AR(1) series shown in Figure 25.5. The first 50 observations represent the in-control process state. Under this in-control state, the observations were generated from an AR(1) process with $\phi_1 = 0.75$, $\mu = 0$ and $\sigma_\varepsilon = 1$. Starting at $t = 51$, a mean shift of size $2\sigma_\varepsilon$ or, equivalently, $0.875\sigma_y$ was introduced. Incidentally, for this example, model estimation using the first 50 observations shows that $\hat{\phi}_1 = 0.746 \approx 0.75$, $\hat{\mu} = 0.0$, and $\hat{\sigma}_\varepsilon = 0.98$.

The corresponding Shewhart chart on residuals detected the shift at $t = 90$ (see Figure 25.6a). A CUSUM on residuals for the series in Figure 25.5 was constructed, using the optimal CUSUMR parameters when $\phi_1 = 0.75$ suggested by Runger *et al*.[12] From Figure 25.6b, we can see that the corresponding CUSUMR detected the shift at $t = 75$.

For an AR(1) series, the autocovariance function is given by

$$\gamma(0) = \frac{1}{1-\phi_1^2}\sigma_\varepsilon^2,$$

$$\gamma(1) = \frac{\phi_1}{1-\phi_1^2}\sigma_\varepsilon^2,$$

$$\gamma(h) = \phi_1\gamma(h-1), \quad h \geq 2.$$

From here, it is easy to calculate the quantity

$$\upsilon_{n-j}^2 = \sum_{|h|<n-j}\left(1 - \frac{|h|}{n-j}\right)\gamma(h), \qquad j = 1, 2, 3, \ldots, n,$$

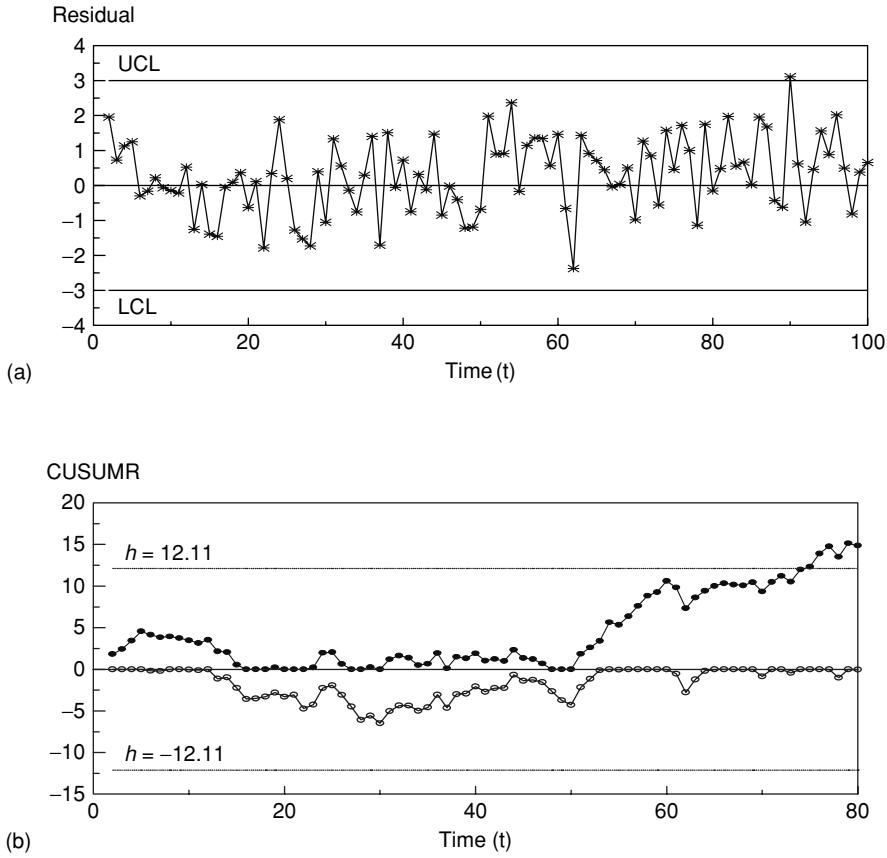


Figure 25.6 Shewhart and CUSUM ($k = 0.125$) charts on residuals for the series in Figure 25.5.

needed for calculating the control limits for the CUSUM. Note that since $\gamma(h) = \gamma(-h)$, then υ^2_{n-j} may also be written as

$$\upsilon^2_{n-j} = \begin{cases} 0, & j = n, \\ \gamma(0), & j = n-1, \\ \gamma(0) + 2\sum_{i=1}^{n-j-1}\left(1 - \frac{i}{n-j}\right)\gamma(i), & j = n-2, n-3, \ldots, 1. \end{cases}$$

The above expression can be implemented using standard spreadsheet packages such as Microsoft Excel and Lotus 1-2-3. For illustration purposes, assume that we want to establish a CUSUM mask that focuses only on the latest 10 observations (i.e. $mw = 10$). This means we only need to calculate the first 10 υ^2_{n-j} values (i.e., for $j = n$, $n-1, \ldots, n-9$). For an AR(1) process with $\phi_1 = 0.75$, $\mu = 0$, and $\sigma_\varepsilon = 1$ these values are given in Table 25.6. Once the υ^2_{n-j} values are available, we just need to specify

Table 25.6 CUSUM mask for an AR(1) process with $\phi_1 = 0.75$, $\mu = 0$, and $\sigma_\varepsilon = 1$ (mw = 10).

$n-j$	$\gamma\,(n-j)$	υ^2_{n-j}	Lower arm	Upper arm
9	0.172	10.823	−29.303	29.303
8	0.229	10.362	−27.032	27.032
7	0.305	9.829	−24.628	24.628
6	0.407	9.209	−22.070	22.070
5	0.542	8.484	−19.338	19.338
4	0.723	7.632	−16.405	16.405
3	0.964	5.429	−13.236	13.236
2	1.286	4.000	−9.783	9.783
1	1.714	2.286	−5.938	5.938
0	2.286	0.000	0.000	0.000

the required constant z^* to establish the CUSUM mask given by equation (25.22) or (25.23). Assuming an ARL_0 of 370 for a two-sided CUSUM scheme (i.e. an ARL_0 of approximately 740 for a one-sided scheme) is required, we can see from Table 25.5 that we need a z^*-value equals to 2.969. The summary of the calculations done using a spreadsheet is given in Table 25.6, while the resulting CUSUM mask is shown in Figure 25.7.

In monitoring the process, we just need to superimpose the vertex of the CUSUM mask in Figure 25.7 on the latest CUSUM value. Figure 25.8 shows how the CUSUM mask in Figure 25.7 detected the change in the mean of the process at $t = 60$, or 10 observations after the change was introduce. Since our mask is based only on $mw = 10$, it is not necessary for us to plot all the historical CUSUM values. The same result will be obtained if we are only maintaining the latest 10 observations.

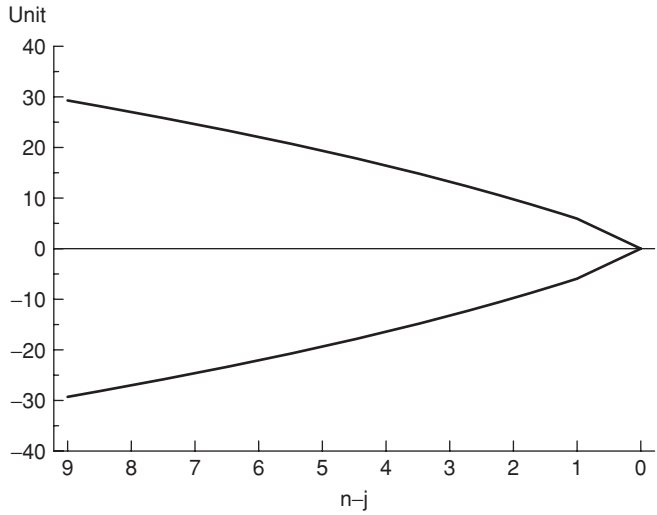


Figure 25.7 Plot of the CUSUM mask in Table 25.6.

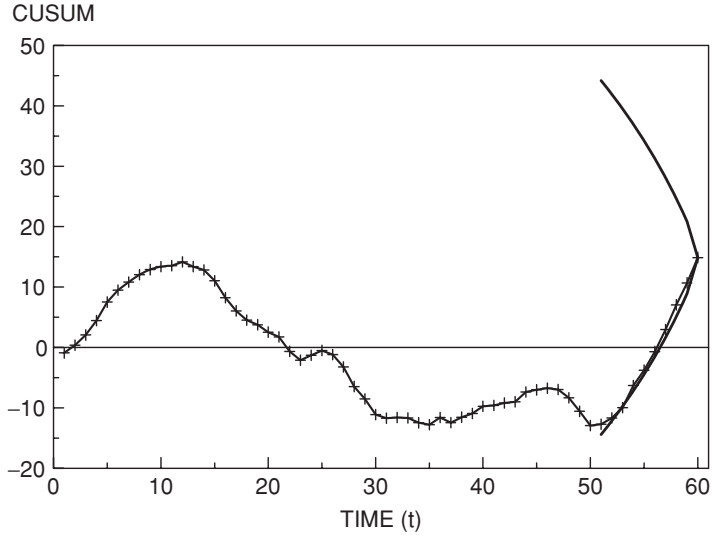


Figure 25.8 The CUSUM chart for from Atienza *et al.*[28]

25.5 CONCLUSION

A CUSUM scheme that utilizes the BCUSUM parabolic mask produces superior ARL properties that rival the performance of a CUSUM using the existing masks (i.e. V, semiparabolic, and snub-nosed). The proposed parabolic mask is theoretically founded. One of its important characteristics is that the user is not expected to specify the size of shift it is most desired to detect. Compared to the tabular form of CUSUM, which involves formulas that may intimidate users with a non-mathematical background, the parabolic CUSUM is easier to implement and understand. One can easily see a decrease or increase the in mean using the plot of sums of deviations from the target or nominal. The use of CUSUM with a parabolic mask can become a lot easier when implemented using a computer.

The Page CUSUM scheme results in powerful procedures for detecting shifts in the process mean of an i.i.d series of process measurements. Under certain conditions, it has been proven optimal in the sense that, among all schemes with the same rate of false alarms, the CUSUM provides the best possible sensitivity. Unfortunately, its performance deteriorates when the process observations are autocorrelated. The popular way of monitoring changes in the mean of an autocorrelated series is to apply the CUSUM or the other classical SPC techniques on the residuals of a chosen time-series model that best explains the dynamics of the process observations. Using the symmetric relation of the CUSUM and BCUSUM together with our knowledge of the distribution of means from an ARMA process, a CUSUM scheme that directly applies the correlated measurements in process monitoring is derived. The performance of the proposed CUSUM scheme for autocorrelated measurements is very competitive in comparison with the SCR and CUSUMR.

REFERENCES

1. Lucas, J.M. (1973) A modified V-mask control scheme. *Technometrics*, 15, 833–847.
2. Bissel, A.F. (1979) A semi-parabolic mask for CUSUM charts. *The Statistician*, 28, 1–7.
3. Rowlands, J., Nix, B., Abdollahian, M. and Kemp, K. (1982) Snub-nose V-mask control schemes. *The Statistician*, 31, 1–10.
4. Barnard, G.A. (1959) Control charts and stochastic processes. *Journal of the Royal Statistical Society* B, 21, 239–257.
5. Van Dobben de Bruyn, C.S. (1968) *Cumulative Sum Tests: Theory and Practice*. London: Griffin.
6. Wiklund, S.J. (1997) Parabolic CUSUM control charts. *Communications in Statistics: Computation and Simulation*, 26, 107–123.
7. Lorden, G. (1971) Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42, 1897–1908.
8. Harris, T.J. and Ross, W.H. (1991) Statistical process control procedures for correlated observations. *Canadian Journal of Chemical Engineering*, 69, 48–57.
9. Alwan, L.C. (1992) Effects of autocorrelation on control chart performance. *Communications in Statistics: Theory and Methods*, 21, 1025–1049.
10. Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994) *Time Series Analysis, Forecasting and Control*. Englewood Cliffs, NJ: Prentice Hall.
11. Wardell, D.G., Moskowitz, H. and Plante, R.D. (1994) Run-length distributions of special-cause control charts for correlated processes (with discussions). *Technometrics*, 36, 3–27.
12. Runger, G.C., Willemain, T.R. and Prabhu, S. (1995) Average run lengths for CUSUM control charts applied to residuals. *Communications in Statistics: Theory and Methods*, 24, 273–282.
13. Yashchin, E. (1993) Performance of CUSUM control schemes for serially correlated observations. *Technometrics*, 35, 37–52.
14. Wardell, D.G., Moskowitz, H. and Plante, R.D. (1995) Letter to the Editor. *Technometrics*, 37, 243–245.
15. Atienza, O.O., Tang, L.C. and Ang, B.W. (2000) A uniformly most powerful cumulative sum scheme based on symmetry. *The Statistician*, 49, 209–217.
16. Harrison, P.J. and Davies, O.L. The use of cumulative (CUSUM) techniques for the control of routine forecasts of product demand. *Operations Research*, 12, 325–333.
17. Gilchrist, W. (1976) *Statistical Forecasting*. New York: John Wiley & Sons, Inc.
18. Makridakis, S. and Wheelwright, S.C. (1987) *Forecasting Methods for Management*, 4th edition. New York: John Wiley & Sons, Inc.
19. Gardner E.S. (1983) Automatic monitoring of forecast errors. *Journal of Forecasting*, 4, 1–28.
20. Atienza, O.O., Ang, B.W. and Tang, L.C. (1997) Statistical process control and forecasting. *International Journal of Quality Scieince*, 2, 37–51.
21. Adams, B.M., Lowry, C. and Woodall, W.H. (1992) The use (and misuse) of false alarm probabilities in control chart design. In H.J. Lenz, G.B. Wetherill and P.T. Wilrich (eds), *Frontiers in Statistical Quality Control 4*, Heidelberg: Physica-Verlag.
22. Tsay, R.S. (1988) Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7, 1–20.
23. Woodall, W.H. and Adams, B.M. (1993) The statistical design of CUSUM charts. *Quality Engineering*, 5, 559–570.
24. IMSL (1991) *User's Manual Statistical Library*, Version 2. Houston, TX: IMSL.
25. Lucas, J.M. (1982) Combined Shewhart-cusum quality control schemes. *Journal of Quality Technology*, 14, 51–59.
26. Brockwell, P.J. and Davis, R.A. (1987) *Time Series Analysis: Theory and Methods*. New York: Springer-Verlag.
27. Adams, B.M. and Tseng, I.T. (1998) Robustness of forecast-based monitoring schemes. *Journal of Quality Technology*, 30, 328–339.
28. Atienza, O.O., Tang, L.C. and Ang, B.W. (2002) A CUSUM scheme for autocorrelated observations. *Journal of Quality Technology*, 34, 187–199.

Index

1.5 sigma shift, 25, 349, 365
2^k factorial design, 15, 239–240, 242, 245–246, 299

A

acceptance sampling, 269
adaptive designs, 264
affinity diagram, 47
akaike information criterion, 191
alternative hypothesis, 154, 175, 184, 385
analysis of variance (ANOVA), 14, 47, 50, 53, 57, 82–83, 242
Anderson Darling tests, 134, 162
algorithmic cusum, 372
average run length, 326, 330, 344, 363, 372, 379
array,
  inner and outer, 239, 251, 276, 288, 290
  orthogonal, 251, 258, 260–261, 265, 276–277, 280, 283, 285
  Taguchi, 285
assignable cause, 75–76, 372
attributes:
  in Six Sigma implementation, 25
attribute data, 211
attribute Shewhart charts, 325
autocorrelation, 243, 343–348, 351–354, 365–366, 382, 391, 394–395
autocovariance, 345, 391, 392, 394, 401
autogressive moving average model, 382
autoregressive integrated moving average, 366
availability, data, 55, 258
average run length, 326, 330, 344, 363, 372, 379

B

balanced:
  assessment of Taguchi methods, 292
  design, 261
  view, 27
baseline:
  defect rate, 32
  measurement, 10, 11, 16
  performance, 131
batch means control chart, 354
bell-shaped curve, 212, 216
benchmark, 225
benchmarking, 47
bernoulli distribution, 182, 183
best-fit distribution, 136, 138–141, 143, 145–149
bias:
  due to scaling, 179
  of measuring Cp(u, v), 128
bias-corrected percentile bootstrap (BCPS), 197, 208
biased estimator:
  population variance, 77
binomial distribution, 179–180, 182, 214
blocking:
  in DOE, 53, 239, 251
blocks:
  replicated runs, 253–255
bottleneck, 93
box plot, 107, 116, 119–128
box-and-whisker, 114
Box-Cox:
  transformation, 119, 120–122, 125–126, 136, 144
  plot, 136, 143, 149

Six Sigma: Advance Tools for Black Belts and Master Black Belts L. C. Tang, T. N. Goh, H. S. Yam and T. Yoap

© 2006 John Wiley & Sons, Ltd

Box-Jenkins ARMA, 382
brainstorming, 12, 47, 87, 89, 91, 94–95
breakthrough:
  strategy, 5, 24, 27–28, 41, 48
  results, 50, 153, 287
  technology, 283
business-improvement, 51, 55

C

calibration:
  of progress, 21
  tests, 226
capability:
  analysis, 53
  assessments, 46
  capability (Monte Carlo simulation), 145
  measurement systems, 32, 45
  of process (Non-normal data), 131, 135
  process, 5, 73, 73, 79, 107–108, 127, 129, 132, 149, 195, 198, 227
  process (short-terms), 140
  process capability index (CPI), 107–149
product design, 57
  Six Sigma, 36
  SPC, 23
  Wrights process capability, 113
categorical:
  data, 171, 173
  scale, 172
  variables in logistic regression (addition and deletion), 192
  variables in logistic regression (multiple), 188
  variables in logistic regression (single), 186
causal:
  hypothesis, 6
  relationship, 172
cause-and-effect diagram:
  analysis, 47
  diagram, 12, 85
  matrix, 13, 101
  relationship, 90, 94, 96, 271, 277, 296–297
CCC, *see* cumulative conformance counts
censored:
  time, 46
  observations, 163
certification:
  QS, 21
  Six Sigma, 27, 35
champion:
  deployment, project, senior, 8
  training, 9, 22, 24
  Six Sigma, 32–34, 37
charts:
  attribute Shewhart, 325
  control, 47, 51, 53, 75, 211–212, 219
  cumulative conformance counts (CCC), 325
  cusum control, 354, 366
  cusum control (with fast response), 371
  EWMA, 212–213, 219, 354, 366
  Gantt, 23
  MH charts, 347
  pareto, 47
  run, 53, 57
coded:
  values in DOE, 240, 263
coefficient:
  autocorrelation, 391
  Pearson correlation, 139
  skewness, 147
confidence interval:
  bootstrap, 208
  differences between proportions, 179
  for acceleration factors, 233
  for C_p, 199
  for C_{pk}, 197, 203
  for percentiles, 138
  in DMAIC, 53, 54
  of odds ratio, 181
  of relative risks, 179
conforming:
  cumulative counts, *see* charts, CCC
  fraction (percentage) nonconforming, 114, 195–196, 207, 325, 335
  proportion nonconforming using Burr's distribution, 108
confounding and aliasing, 247, 249–250, 252, 254, 284, 298–299
constraint:
  costs, 44, 76
  in experimental designs, 254, 297
  production processes, 41
  response in dual responses, 308
  safety, 42
  theory of, 93
consumer-oriented:
  packaging, 292
contingency tables, 171
contour plot, 198
control charts, *see* charts
control phase, 6, 8, 21, 51, 54, 63, 131, 323
correlation and regression analysis, 53
costs of poor quality, 4
covariance, 309, 345, 350, 391–392, 394, 401
covariance matrix, 345, 350
C_p, 107–149
C_{pk}, 46, 107–149
C_{pl}, 109
C_{pm}, 144
C_{pmk}, 108
C_p (u, v), 108

C_{pu}, 108
C_s (Wright's process capability index), 113
Cramer von Mises tests, 161
cross-classifies:
  contingency table, 176, 177
cross-functional:
  projects, 8
  teams, 42
Current Reality Tree (CRT), 93–95
CTQ, 5, 21, 25, 27–28, 32–33, 42, 45, 50
cumulant, 167, 169
cumulative distribution function (CDF):
  lognormal and Weibull, 114
  standard normal, 134
curvature, 86, 240, 290
customer requirements, 24, 48
customer satisfaction, 38780, 21–22, 24, 27, 31–33, 36, 44, 50–51, 73, 94, 172, 272
customer-centered, 49
CUSUM:
  backward cusum (BCUSUM), 382, 404
  charts, 353–366, 371–379
  charts on residuals, 364–366
  CUSUM chart, 212, 213, 219, 364
  CUSUMR, 379, 395–397, 401
  FIR CUSUM, 374–379
  multivariate scheme, 354
  S-CUSUM, 390
  tabular CUSUM, 371–379
  V-mask CUSUM, 371–379
  zero-start CUSUM, 374

D

data transformation, 211, 212, 214, 216
data-driven, 17, 33, 38, 104, 153
defect, 4, 5, 10, 21–25, 27, 32, 50–51, 55–56, 73–74, 131, 196, 206, 212, 221, 270–272
defective, 24, 74, 196, 221, 270–271
defects per million opportunities, 4, 22, 131
deployment, 3, 7, 8, 10, 21–23, 33–34, 41–42, 48, 50–52, 54–55, 57, 61–63, 91, 94, 292
design of experiments, 21, 34, 50, 57, 101, 155, 239, 242, 244, 250, 252, 254, 257–258, 261, 268, 271–273, 275–276, 280, 283, 288, 297–298
design resolution, 250
design-oriented, 48
destructive tests, 264, 303
deterministic service times, 66
deviance statistics, 186
distribution-free, 108, 109, 113, 127, 154
DMADOV, 48
DMADV, 41
DMAIC, 3–7, 20–23, 26, 31–32, 34, 36, 38, 41, 44, 47, 49, 56, 58, 74–75, 86, 101, 131
dpmo, 22, 24, 131–132
dpo, 112–113, 198
DRSALG algorithm, 310
DR2 algorithm, 310

E

error:
  type I, 162, 326–327, 373
  type II, 373
estimator, 73, 77–79, 108, 110–111, 156, 174, 327–330, 334, 338–339
evolutionary operation (EVOP), 271, 277, 289, 290
empirical distribution function (EDF): *see* goodness-of-fit tests, Empirical Distribution Function based approach
EWMA:
  charts, *see* charts, EWMA
  control schemes, 212, 221
  MCEWMA, 354
  multivariate control charts, 354
exponential distribution, 45, 137, 139

F

face-centered design (FCD), 319
factorial design, 15, 239–242, 244–250, 257, 264, 276, 290, 299
failure mode and effect analysis (FMEA), 42, 45, 47, 50, 53, 94, 101–104
fishbone diagram, 12, 85–91, 103
fisher's cumulant tests, 167–169
fisher's methodology:
  design of experiments, 271
flip-chart, 88
flowchart, 96, 226
fold-over designs, 298, 302–303
fractional factorials, 47, 259, 284, 288, 298
framework, 19–23, 31–38, 41–42, 48–50, 54, 56, 58, 154, 225, 274, 311, 326
future reality tree (FRT), 93–94, 96–100, 102
fuzzy approach for dual response, 309

G

gage r&r, 47
gamma:
  function, 78, 196
  distributions, 112
Gantt charts, 43, 47
Gaussian:
  in DOE, 246, 259, 261–262
  process, 345, 351
generators:
  stationary time series, 343
genetic algorithms, 13
geometric distribution, 211–216, 219, 221, 326

goal:
current reality tree, 94–96
of Six Sigma, 4–17, 24–33
programming, 53, 307, 309, 312
goalpost loss function, 250
goodness-of-fit tests (GOF):
general, 138–139, 153–169
principles, 154
Pearson Chi Square, 155
Empirical Distribution Function based approach (EDF), 157
Kolmogorov-Smirnov, 158–161
Cramer von-Mises, 161
Anderson Darling, 134–162
regression based approaches, 163
probability plotting, 163
Shapiro Wilk, 165
Fisher's Cumulant, 167–169
greco-latin designs, 261
green belt, 21, 25, 32, 34

H

heuristic, 53–54, 57, 108, 110
high-yield process, 128, 212, 325
histogram, 45, 132–135, 140–143, 146–148, 215–216, 219
homogeneous:
batch processes, 197
subgroups, 76
Hosmer-Lemeshow, 186
house of quality (HOQ), 13, 42–43
hypothesis:
causal hypothesis, 6
difference betweeen likelihood ratios, 191
testing, 50, 53–54, 154
of Anderson Darling, 134
tests for GOF, *see* goodness-of-fit tests
for contingency table, 173
for contingency table with ordinal

I

information, 175
for logistic regression, 183–184
for detecting outlier and level shifts, 357–358
identify-design-optimize-validate (IDOV), 22–23, 41–43, 45, 47
IMSL:
FORTRAN library, 366
statistical library, 347, 388
subroutines, 203
independence:
in contingency tables, 174–178
in SPC control charts, 344, 366, 382
Negatives in CRT, 94
of residuals, 243
inferences, 138, 172, 181
innovation, 27–28, 48, 347, 354, 356, 362, 366
innovation algorithm, 366
input-output, 24, 271, 277, 288
inspections, 326, 330
interaction effects, 191, 234, 239–257, 264, 271–288, 298–306
ishikawa diagram, 12, 62, 85
ISO, 21, 268

J

Johnson transformation, 111–113, 127

K

kaizen, 64
Kano analysis, 53, 57
key variables, 6
knapsack, 54
Kolmogorov-Smirnov test, 158–161
Key Process Input Variables (KPIVs), 32, 44–45, 171–172
Key Process Output Variables (KPOVs), 32, 44–47, 171–172
kurtosis, 136–137, 167, 169

L

Lagrangian multiplier approach, 308–309
Laplace, 137
lean manufacturing:
manufacturing, 51
design, 258, 262, 297–305
least squares, 137–138, 171
level of significance, 161–162, 167, 199
likelihood:
maximum likelihood estimation (MLE), 111, 137–138
likelihood ratio, 174, 184–186, 191–192, 197, 199, 354, 381, 384–385, 385
linear regression, 57, 155, 159, 171, 188, 191, 357
linearization, 164
location parameter, 229–233, 272
logic-based tool, 94
logistic regression, 181–192
logit link, 181–182, 186–187
log-likelihood, 128
loglogistic:
distribution, 140, 228–232
three-parameter, 137, 139
lognormal:
system, 111–112, 112
distribution, 114–121, 127–129, 143–149
three-parameter, 137, 139
loss function, 250–251, 273, 279, 287, 290, 309

M
main effect of a factor, 241
maintainability, 43
maintenance, 7, 27, 28, 43, 51, 55, 356
mapping, 53, 57, 85, 90–91
Markov, 60, 379
master black belt, 7–9, 21, 27, 32–34, 51
maturity, 291
mean squares, 83
meanshift, 384
metaheuristics, 54
monte carlo simulation, 34, 45, 47, 53, 57, 79, 128, 132, 145, 147, 160–161, 347, 358, 365, 379, 388
multiple regression, 47, 189
multiplier, 308, 309
multiresponse, 309
multivariate, 155, 345, 354–355
multivariate ewma control chart, 354

N
near-zero-defect, 212
negative binomial distribution, 214
negative correlation, 178
nested design, 73, 79–82, 244–245
Newton-Raphson, 138
noise factors, 13, 239, 245, 251–252, 254
nominal value, 275, 279
nonnormal data:
  nonnormality, 107–116, 120–129, 219, 221
nonparametric approach, 134, 154, 163, 195, 197, 207, 216
normal distribution, 77, 112–118, 127–136, 153–157, 161–169, 184, 186, 211–219, 228
normal probability plots, 165
normality test, 133–135, 141–148, 219
normalized, 44
null hypothesis, 134–155, 155–157, 160–161, 165, 167, 169, 173–178, 184–188, 357–358

O
optimal, 14–15, 23, 44, 54, 111, 128, 197, 251, 270, 278, 289–290, 307, 310, 318, 379, 401, 404
optimality, 23, 289–290, 310
optimization, 24, 28, 44, 48, 50–54, 57, 93–94, 138, 153, 203, 240, 257, 272–274, 278, 284, 287, 289–290, 307–318
ordinal information, 172, 175, 177–178
orthogonal array, 258, 260–261, 264, 276–277, 280, 283–285, 288
outlier, 58–59, 116, 163, 354–358, 362–363, 366

P
p chart, 212, 270, 325
parameter estimation, 137, 185, 326, 330
parametric, 134, 138, 154, 163
Pareto analysis, 264, 302
Pareto chart, 13, 47
Pearson correlation, 139
Pearson chi-square tests, 155
Pollaczek-Khintchine, 65
population:
  population mean, 77–78, 156, 161–162
  population variance, 76–77, 253
positive correlation, 44
P_{pk}, 140, 144–145
predictive, 45, 188, 191–192
predictor, 187, 189, 192, 243
probability plot, 109, 113, 119–128, 138–139, 143–144, 163–166, 228–229
process capability, 5, 46, 73, 79, 107–114, 116, 120, 122–136, 138, 140–149, 195–198, 206, 269–270, 278
process capability index (PCI), 107–115, 127–129
process capability analysis (PCA), 129–133, 142, 148–149
process characterization, 28, 79
process design, 7, 273, 288, 297
process improvement, 8, 32–33, 37, 328, 337–338
process mapping, 53, 57
process-focused, 49, 52
process-improvement, 8, 51
process-monitoring, 211, 213
producibility, 28, 75, 226
proportion defective, 270

Q
q-charts, 221
Quality function deployment (QFD), 21, 34, 42–48, 50–53, 57, 91
q-transformation, 212, 217, 219
quality and productivity, 269
quality characteristics, 32, 76, 211, 212, 214, 216, 221
quality engineering, 19, 28, 58, 129, 280, 287–289, 297, 306, 343–344
quality improvement, 52
quantile, 164, 165

R
random:
  sampling, 79
  variable, 45, 77–78, 115, 165, 171, 179, 182, 186, 326–327, 347, 355, 359, 366, 381
  randomization, 241, 245, 290
range, 8, 10, 26, 49, 51, 62, 81, 95–98, 100–101, 107, 111–112, 128–129, 154–155, 175, 195, 197–198, 203, 205, 234
rank, 9, 13, 112, 164–166, 197, 208, 281, 283

rank-ordering, 165
rational subgroups, 76
regression:
  analysis, 13, 50, 53, 156–163, 223, 225, 357
  model, 155, 181–192
reliability engineering, 36, 38, 221, 230, 257, 306
repeatability, 75, 226, 282
replicate, 15, 79, 116, 203, 208, 239, 246, 251–255, 299, 307, 311, 315, 318–319, 375, 379, 388, 395
replication, 45, 290, 298, 304
  unreplication, 399
reproducibility, 75, 226
residual, 111, 138, 155–162, 165–166, 174–178, 185–186, 231–232, 239, 243–244, 344, 347, 353–355, 364–366, 382, 395, 401–402, 404
residual analysis, 231, 239, 243
residual control charts, 354
risk, 42, 44–45, 53, 57, 153–154, 173, 178–180, 189, 327, 331, 373
roadmap, 3–5, 8, 41–57
robust process, 57
robustization, 280
robustness, 14, 23, 61, 113, 129, 154, 278–280, 290–291
Root cause analysis (RCA), 62, 101–104
run chart, 53, 57

S

safety, 42, 258
sample:
  size, 46–47, 110, 113, 127, 129, 135, 158, 196, 198, 208, 226, 325–329, 334, 336, 338, 357
  standard deviation, 77, 79, 252–253
  variance, 77, 196, 252
sampling distribution, 113, 179, 181, 196, 358, 363
saturation, 224, 226–228, 230
scatterplots, 189
screening, 47, 62–64, 141, 246, 263–264, 289–290
screening experiment, 47, 263–264
sensitivity analysis, 45–47, 52–54, 59–64, 155, 175, 212, 214, 219, 221, 225, 235, 278, 308, 313, 315–316, 318, 320, 328, 332, 343, 346–347, 348
sequential experimentation, 258, 290, 298
sequential sampling, 331
Shapiro Wilk tests, 165
Shewhart control chart (SCC), 216, 220, 344, 346, 348, 354, 371, 379
signal-to-noise, 276, 284, 307
six sigma:
  organization, 7
  process improvement, 32
  roadmap, 4–5, 52, 54
skewness, 107–115, 128–129, 135–137, 141, 143, 146–148, 163, 167, 169, 217, 219
solver, 207, 307–308, 310, 320
specification limit, 132–133, 135, 141–142, 145, 196–198, 201, 226, 250, 279, 394
standard deviation, 76–77, 79, 108, 198, 252–253, 307–310, 312
standard errors, 233
standard normal distribution, 113, 134, 169, 214
stationary process, 393
Statistical process control (SPC), 21–23, 50, 53, 57, 269, 271, 344, 353–354, 358, 360, 362–363, 382, 384, 386, 391, 393, 404
statistical control, 32, 75, 205, 221, 328, 330, 355
statistical hypothesis, *see* hypothesis
statistical inference, 172, 178, 181, 257, 290
statistical model, 153–154, 171–172
statistical tests, 132, 174
subgroup, 76, 243, 253–254
suboptimization, 28, 94
SWOT Analysis, 31–38
symmetric, 83, 136, 167, 213, 387–389, 404

T

t-test, 290
tabular, 173, 371–376, 378–379, 381, 404
tabular cusum, 372, 374–375
taguchi methods, 33, 257, 261, 264, 267–268, 272–292
taguchi orthogonal array, 260–261
test statistic, 155, 161–162, 167, 169, 216, 219, 354, 357, 385, 391
Theory of constraints (TOC), 93, 104
three-parameter, 137, 139–140, 143, 145–148
three-sigma, 212
threshold, 131, 137, 139–140, 144–148, 371–372, 374, 386
time series graph, 53
time series model, 353–354, 354
time-censored, 46
tolerance:
  bilateral tolerance limits, 196
  design, 54
  graphical tolerance box, 197
  settings in Optimize phase, 45–46
  single tolerance limit, 114, 196
  tolerance intervals, 109–128
  via Monte Carlo simulation, 45
toolbox, 85–86
total quality management (TQM), 4, 33, 35–36
transformation, 111, 115, 127

transient, 37, 224,
treatment factors, 79, 239, 251–252
two-parameter exponential, *see* exponential distribution
type I error, *see* error
type II error, *see* error

U

unbiased estimator:
  error variance, 231
  p, 327
  population standard deviation, 79
  population variance, 77–78
  s/c_4, 79
  variability, 73
uniformity, 24, 283–284
unimodal, 216
unreplicated, *see* replication
unstable, 253
upstream, 22, 23, 41–42, 269

V

value-added, 59, 95
variability, 5, 57–58, 73, 75–77, 113–114, 116, 127, 200–202, 206, 251, 278, 280, 307
variance components, 73, 82–83
variance-covariance, 309
V-mask, 371–382, 388–390, 393
voice of the customer, 42

W

wafer, 74, 81–83
Weibull distribution:
  distribution, 114–116, 123–124, 127, 137
  three-parameter, 137, 139
weighted variance, 108, 110, 127
Wright's process capability index, 113, 127

Z

zero defects, 5, 21, 212, 221
z-statistic, 184, 186, 391